AF/1645

PATENT
Attorney Docket No. AMGEN-08341

123

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:	Marek Z. Kubin *et al.*		
Serial No.:	09/667,859	Group No.:	1645
Filed:	09/20/2000	Examiner:	B. Li
Entitled:	**NK Cell Activation Inducing Ligand (NAIL) DNA And Polypeptides, And Uses Thereof**		

TRANSMITTAL OF APPLICANT'S REPLY BRIEF (PATENT APPLICATION - 37 CFR § 192)

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

CERTIFICATE OF MAILING UNDER 37 C.F.R. § 1.8(a)(1)(i)(A)

I hereby certify that this correspondence (along with any referred to as being attached or enclosed) is, on the date shown below, being deposited with the U.S. Postal Service with sufficient postage as first class mail in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450

Dated: March 8, 2004 By: Mary Ellen Waite

Sir or Madam:

Applicant submits, in triplicate, the REPLY BRIEF to the Examiner's Answer (Paper No. 15) mailed January 6, 2004, in the above application. Applicants believe no fee is required, but if the Commissioner deems otherwise, the Commissioner is hereby authorized to charge Deposit Account NO. 02-190 any fees associated with this communication.

Dated: March 8, 2004

J. Mitchell Jones
Registration No. 44,174
MEDLEN & CARROLL, LLP
101 Howard Street, Suite 350
San Francisco, California 94105
608/218-6900

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:	Marek Z. Kubin and Raymond G. Goodwin		
Serial No.:	09/667,859	Group No.:	1645
Filed:	09/20/2000	Examiner:	B. Li
Entitled:	**NK Cell Activation Inducing Ligand (NAIL) DNA and Polypeptides, and Uses Thereof**		

APPELLANTS' REPLY BRIEF

Mail Stop Appeal Brief - Patents
Commissioner for Patents and Trademarks
P.O. Box 1450
Alexandria, VA 22313-1450

CERTIFICATE OF MAILING UNDER 37 C.F.R. § 1.8(a)(1)(i)(A)

I hereby certify that this correspondence (along with any referred to as being attached or enclosed) is, on the date shown below, being deposited with the U.S. Postal Service with sufficient postage as first class mail in an envelope addressed to: Mail Stop Appeal Brief - Patents, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

Dated: March 8, 2004

By: _____
Mary Ellen White

Sir:

This Brief is in reply to the Examiner's Answer (Paper No. 15) mailed January 6, 2004.

It is not believed that any fees are necessary for this reply. However, if any fees are necessary, the Examiner is hereby authorized to charge Deposit Account No. 08-1290 the fee associated with this extension and any other fees associated with this communication. Please reference Attorney Docket No.: AMGEN-08341 when charging the Attorney Deposit Account..

This Brief is transmitted in triplicate. [37 C.F.R. § 1.192(a)].

ARGUMENT

The Office's acceptance of the statements of the real party in interest, status of claims, status of amendments after final, summary of invention, and issues, and grouping of the claims is appreciated. However, the Appellants respectfully submit that the Office is mistaken with respect to the lack of a statement of related appeals and interferences. Appellants included the required statement in their Appeal brief. The Office is respectfully directed to Page 3 of the Appeal Brief, where Appellants state: "There are no related appeals or interferences known to Appellants, Appellants' legal representative, or the Assignee." Appellants also thank the Office for withdrawing the rejection of the claims 73-84 under 35 U.S.C. 112, second paragraph.

Below, Appellants specifically address the following issues from the initial Appeal Brief:

1. Whether Claims 73, 74, 80, and 84-89 are enabled under 35 U.S.C. §112, first paragraph;

2. Whether Claims 73, 74, 80, and 84-89 are supported by an adequate written description under 35 U.S.C. §112, first paragraph; and

3. Whether Claims 73-78 and 80-89 are patentable under 35 U.S.C. § 103(a) over Valiante et al. (U.S. Pat. No. 5,688,690), Sambrook et al. (Molecular Cloning - A Laboratory Manual, 2nd Edition, Cold Spring Harbor, N.Y. 1989, pp. 2.43 - 2.84) and Porunellor et al. (J. Immunol. (1993) 151:5328 - 5337).

1. Claims 73, 74, 80, and 84-89 are enabled under 35 U.S.C. §112, first paragraph

Claims 73, 74, 80 and 84-89 remain rejected as allegedly being non-enabled. Appellants note with appreciation the Office's removal of this rejection with respect to Claims 75-78 and 81-83. Appellants' maintain that the Office's rejection of the claims as non-enabled for the

following reasons: a) the Office's reliance on Struffey et al. and Proudfoot et al. is scientifically invalid and irrelevant to the Claims at issue; b) the Office's analysis under In re Wands is both new and misguided;

a) The Office's reliance on Stuffey et al. and Proudfoot et al. is scientifically invalid and irrelevant to the Claims at issue

As presented in Appellants' Appeal Brief, the standard to be applied in assessing enablement is whether the experimentation needed to practice the claimed invention is undue or unreasonable. *See* TRAINING MATERIALS FOR EXAMINING PATENT APPLICATIONS WITH RESPECT TO 35 U.S.C. SECTION 112, FIRST PARAGRAPH-ENABLEMENT CHEMICAL/BIOTECHNICAL APPLICATIONS, *citing In re Wands*, 8 USPQ2d 1400, 1404 (Fed. Cir. 1988). When applying this standard, the burden is on the Office to make a *prima facie* case of non-enablement that is well grounded in scientific reasoning or evidence. *See In re Wright*, 27 USPQ2d 1510 (Fed. Cir. 1993); *See also* MPEP §706.03 and §2164.04. This is because without a reason to doubt the truth of the statements made in the patent application, the application must be considered enabling (*Wright,* 27 USPQ2d at 1513). The Office's reply makes no reference to these standards and fails to address the Appellants' arguments regarding the scientifically inappropriate analogy drawn by the Office between the claimed invention and several prior art references.

The Office's main attack on the claims relies on the teachings of three references: Robin et al., Struffey et al., and Proudfoot et al. In paragraph 4 of the Examiner's Answer, the Office admits that the isolation of clones is well known in the art. The Office states that highly homologous proteins can be "functionally different molecules." In paragraph 5, the Examiner states that Robin et al. teaches that even though human chemokines MIP-2a, MIP-2b, and

GRO/MGSA have a homology of 87%, they are functionally different molecules. In paragraph 6, the Examiner states that Struffey et al. and Proudfoot et al. show that mutation of molecules is unpredictable because a single amino acid mutation can change a proteins biological function and turn it into a patentably distinct subject.

This reasoning is irrelevant to the claims at issue and is not based on sound scientific reasoning. First, the Office has failed to take into account the limitations of the Claims that require that the claimed sequences encode a polypeptide that binds CD48. The function must be preserved. The Office's arguments are irrelevant because the references cited by the Examiner refer to a situation where function is not conserved.

Second, the Office has failed to establish any scientific reason as to why proteins encoded by the claimed sequences are analogous to the proteins in the cited references and would be expected to have similar properties. Instead of addressing this issue, the Office has merely argued that "because NAIL polypeptide is a protein, it inherently has all the characteristics of a protein, such as random mutation of amino acid(s) can be unpredictable for maintaining its original biological function as evidenced by Struffey et al. (Eur. J. Immunol. 19989, Vol. 28 pp. 1262-1271) and Proudfoot et al. (US Patent 6,159,711A) discussed supra." (Examiner's Answer, paragraph 20). This argument is not based on sound scientific reasoning because it focuses on two isolated references. The Office's reasoning ignores the many references that positively demonstrate that proteins can be mutated and maintain a biological function. This fact, which was well known in the art at the time of the filing of the application, is established by the many references describing the directed evolution of a wide variety of proteins. A small example of these references include the following: 5,514,568 (filed Jan. 19, 1994; issued May 7, 1996); 5,512,463 (filed June 1, 1994; issued April 30, 1996); 5,811,236 (filed Nov. 30, 1995; issued

Sept. 22, 1998); 5,830,721 (filed March 4, 1996; issued Nov. 3, 1998); 5,824,469 (filed Sept. 30, 1994; issued Oct. 20, 1998); Crameri et al., Nat. Biotechnol. 14(3):315-19 (1996); and Moore et al., J. Mol. Biol. 272(3):336-47 (1997). Copies of the references are provided herewith for the Office's convenience at Tabs 1 - 7, respectively.

These references establish unequivocally that it was well known in the art that mutation of a particular protein can and does lead to many proteins with similar functions. Moreover, it was routine in the art at the time of the filing of the present application to screen large numbers of mutants for mutants with a desired function. Thus, the claimed sequences can be mutated and a great number of sequences can be identified that bind to CD48 as claimed. As the Federal Circuit held in *Ajinomoto Co., Inc. v. Archer-Daniels-Midland Co.*, 228 F.3d 1338 (Fed. Cir. 2000), enablement "is determined from the viewpoint of persons of skill in the field of the invention at the time the patent application was filed." Id. at 1345. In affirming the District Court, the Federal Circuit relied on the District Court's finding that:

> According to the record, all of the methods needed to practice the invention were well known to those skilled in the art. Despite the diversity that existing among bacteria, practitioners of this art were prepared to carry out the identification, isolation, recombination, and transformation steps required to practice the full scope of the claims.

Id. See also *Hybritech Inc. v. Monoclonal Antibodies, Inc.*, 802 F.2d 1367, 1384 (1987) (finding claims enabled where methods were known to persons skilled in the art).

As established above, the mutation and screening of proteins for mutants with conserved function was well known in the art at the time the present application was filed. Thus, under *Anijomoto*, the claims are enabled because such methods for using the sequences referenced in the claims were well known.

In addition to what was known in the art and as established in the Appeal Brief, the specification provides extensive guidance for creating and screening mutants. As taught in the

specification, it is straightforward to determine what variations of these nucleotide and amino acid sequences falls within the 80% sequence identity limitation recited in the claims while maintaining the property of binding CD48. Such variations are described in the specification on pages 18-20 and 24-27, and include those differing from native SEQ ID NO:1 due to mutations, restriction digests, ligation to addition sequences, and chemical synthesis, for example; and those differing from SEQ ID NO:2 due to deletions, insertions, substitutions, and fusions, for example. These additional molecules can be generated according to methods described in the specification and methods well known in the art, such as those provided on pages 18-20, 24-27, 25-33, and Example 3, page 67. A computer program for comparing sequence identity is provided on pages 19 (for nucleotide) and 24 (for amino acid) of the specification. Binding of the polypeptide to CD48 can be determined, for example, using the assays described in detail on pages 46 and 47, and equilibrium binding assays described in Example 8, page 71. In addition, the specification describes on pages 21-24 how to generate fragments of SEQ ID NO:2, and how to test for the ability of the fragment to bind to CD48.

This evidence, **which was not addressed in the Examiner's answer**, establishes that the specification teaches in detail how to: 1) make variants of SEQ ID NOs: 1 and 2; 2) calculate the percent identity between SEQ ID NOs: 1 and 2 and the variant sequence; and 3) test the variant sequence to determine if it binds to CD48. Appellants respectfully submit that the Office is required to consider such evidence. The MPEP states that:

> **Office personnel should consider all rebuttal arguments and evidence presented by Appellants**. . . . *In re Beattie*, 974 F.2d 1309, 1313, 24 USPQ2d 1040, 1042-43 (Fed. Cir. 1992). . . . **Office personnel should avoid giving evidence no weight**, except in rare circumstances. *Id. See also In re Alton*, 76 F.3d 1168, 1174-75, 37 USPQ2d 1578, 1582-83 (Fed. Cir. 1996).

Furthermore:

> If a *prima facie* case is made in the first instance, and if the applicant comes forward with a reasonable rebuttal, whether buttressed by experiment, prior art references, or argument, the entire merits of the matter are to be reweighed.[1]

The Office has failed to consider or address the evidence offered by the Appellants. In particular, the Office has failed: 1) to take into account that the claims specify that mutants bind to CD48; 2) to demonstrate why the cited Struffey et al. and Proudfoot et al. references are relevant to the claimed sequences; 3) to establish a *prima facie* of non-enablement that is well grounded in scientific reasoning because of the lack of relevancy of the cited references and the established prior art that teaches mutation of a protein normally results in mutants with conserved function; and 4) to rebut the Appellants arguments that the specification teach how to make and screen for mutants with the desired CD48 binding function. Given these failures to rebut evidence offered by the Appellants, any prima facie case of non-enablement stands rebutted and the rejection should be removed.

b) The Office's analysis under In re Wands is both new and misguided

The Appellants submitted an *In re Wands* analysis in the Appeal Brief. Instead of addressing the Appellants' analysis, the Office has chosen instead to recycle an analysis that was never applied to the currently pending claims. As explained herein, this recycled analysis is not applicable to the pending claims because they contain different limitations than the claims for which the Office's *Wands* analysis was originally presented. Furthermore, the Office has failed to address or rebut any of the arguments made in the Appellants' *Wands* analysis. Thus, those arguments stand unrebutted.

---

[1] *In re Hedges*, 783 F.2d 1038, 1039, 228 USPQ 685, 686 (Fed. Cir. 1986).

i) The Office's failure to present a *Wands* analysis prior to the Appeal

With respect to the failure to ever present a *Wands* analysis for the pending claims, in paragraph 21 of the Examiner's Reply, the Office argues that "the Office Action, paper No. 7 mailed February 27, 2002 had analyzed each of the Wands factors in detail. In response to the Office Action of paper No. 7, Appellants canceled all rejected claims 48-50, 54-57, and 59 and added new claims 73-89 on page No. 8, filed in on July 2, 2002. However, the broad scope of claims 73, 74, 80, 84-89 are still read on the rejected claims 48 and 59. Therefore, Office further explained the Wands factor in paper No. 9., mailed September 20, 2002."

Appellants respectfully submit that the Office is mistaken with respect to whether pending claims "read-on" canceled claims 48-59. The pending claims are substantially different from Claims 48 and 59 in that they include a functional limitation (binding to CD48) and do not include the claim term "fragment". Thus, the Office's earlier arguments are not applicable to the current claims. Moreover, the Office has completely ignored the effect of the changes in their recycled *Wands* analysis. The arguments presented by the Office in paper No. 9 make no reference to *Wands* and cannot be interpreted to be a discussion of the *Wands* factors with respect to the claims. Thus, the Office's *Wands* analysis of the currently pending Claims is new to the Appellants and fails to consider the actual claimed subject matter.

ii) The recycled Wands analysis fails to address evidence presented by Appellants

In presenting the recycled *Wands* analysis, the Office has failed to address the arguments submitted by the Appellants in their Appeal Brief. Thus, those arguments stand unrebutted. For many of the same reasons as discussed above, the Office's Wands analysis is insufficient.

In paragraphs 25-27, the Office purports to address the state of the art and unpredictability of the field. The Office again relies on the Robin et al., Struffey et al., and Proudfoot et al. references to show unpredictability. However, as described in detail above, the references bear no relevance to the claimed sequences, the references are not in accordance with the great majority of references that that show the proteins can be mutated and retain function and the arguments concerning the references fail to take into account the claim limitation requiring CD48 binding. Moreover, the Federal Circuit has specifically held that routine experimentation, such as creating mutants and screening for function in the present application, does not constitute undue experimentation:

> The test [for undue experimentation] is not merely quantitative, since a considerable amount of experimentation is permissible, if it is merely routine, or if the specification in question provides a reasonable amount of guidance with respect to the direction in which the experimentation should proceed to enable the determination of how to practice a desired embodiment of the claimed invention.

Johns Hopkins Univ. v. Cellpro, Inc., 152 F.3d 1342,1360 (Fed. Cir. 1998)(citing *PPG Indus., Inc. v. Guardian Indus. Corp.,* 75 F.3d 1558, 1564, 37 U.S.P.Q.2D (BNA) 1618, 1623 (Fed. Cir. 1996)). Methods of mutation and screening were well known in the art and routine. Thus, as previously stated in Appellants Appeal Brief, the skill in the art is high and methods of making and testing mutants were well known.

In paragraphs 28-30 of the Examiner's Answer, the Office addresses the number of working examples and amount of guidance. In essence, the Office states that the specification teaches methods for isolating SEQ ID NO:1 and SEQ ID NO:3 and that there are no working examples "to

illustrate that any or all nucleic acid molecules having at least 80% homology to the SEQ ID NO:2 that is able to exhibit the same function as the full length of NAIL and the extracellular domain fusion polypeptide of NAIL." Appellants respectfully note that this argument completely fails to address the Appellants' specific citations to the specification (see discussion above) that demonstrate how to mutate and screen for mutants with the desired activity. The Office should not be allowed to simply ignore the evidence offered by the Appellants and simply make unrelated arguments. Indeed, as described above, the MPEP and the case law specifically state that such evidence must be considered. Regardless of the Office's failure to consider this evidence, it establishes that the Appellants did indeed provide detailed guidance with respect to making and testing mutants with 80% homology to SEQ ID NO:2. When combined with the well-developed state of the art, the invention is easily practiced by one of ordinary skill in the art.

In paragraphs 31 and 32, the Office addresses the scope of the claims. The Office states that the claims "are very broad with the claims reciting any or all nucleotide sequence having more than 80% homology that encodes a polypeptide, which is able to bid CD48 molecule." Appellants respectfully note that this statement actually indicates that the claims are limited in scope, i.e., the claimed sequences must be at least 80% homologous and they must encode a protein that binds CD48.

In paragraphs 33-34, the Office addresses nature of the invention. The Office correctly states that the claims are directly to functional variants of SEQ ID NO:2. However, the Office goes on to state that to enable such an invention "the precise molecular structure being able to exhibit the same identical function should be disclosed." Appellants respectfully submit that that conclusory statement does not belong in the *Wands* analysis. Indeed, the purpose of the *Wands* analysis is to examine the claimed invention under a number of factors to reach a conclusion as to enablement.

Furthermore, the Office provides no case law support for what is essentially a conclusion of law - that enablement in this case requires disclosure of a precise molecular structure. Even if this standard was correct, Appellants have met it. In fact, the Appellants have provided such structures (i.e., SEQ ID NO:1, SEQ ID NO:2 and SEQ ID NO:3) and merely claims sequences that are related to those sequences. Thus, precise molecular structures have been recited in the claims.

In paragraphs 35-36 of the Examiner's Answer, the Office purports to address the level of skill in the art. However, instead of addressing the actual level of skill, the Office states that "significant hurdles remain to be overcome for practice the full scope of claimed invention." Again, this is a conclusory statement that bears no relevance to the issue of the level of skill in the art. As previously established in the Appeal Brief, the level of skill in the art is high. This conclusion is supported by the Federal Circuit, which has found that the level of skill in the art of molecular biology is high. *See, e.g., Enzo Biochem, Inc. v. Calgene, Inc.*, 188 F.3d 1362, 1373 (Fed. Cir. 1999)(holding that the level of skill in the art for molecular biology is that of a post-graduate researcher and recognizing the "rather high level of skill in the art possessed by a post-graduate researcher").

As established above, the Office failed to present a *Wands* analysis of the claims at issue and then attempted to recycle arguments made for claims with different limitations into an analysis of the present claims. In doing this, the Office failed to address or rebut the analysis offered by the Appellants in their Appeal Brief. Thus, those arguments stand unrebutted. Even if the Office had attempted rebut the arguments, it is clear that analysis under the Wand's factors supports a finding of enablement as established in Appellants' Appeal Brief.

2. The claims are supported by an adequate written description

Claims 73, 74, 80 and 84-89 remain rejected under 35 U.S.C. 112, first paragraph, as containing subject matter which was not described in the specification in such a way as to reasonably convey to one of skill in the art that the claimed invention was in the possession of the inventor at the time of filing. In maintaining this rejection, the Office has continued to misinterpret the holding of Eli Lilly as it applies to the present claims and has completely failed to rebut the Appellants arguments with respect to the Office's written description guidelines.

a. The Office's *Eli Lilly* argument is unfounded

In paragraph 41 of the Examiner's answer, the Office states that *Eli Lilly* (*Regents of the University of California v. Eli Lilly and Co.*, 119 F.3d 1559 (Fed. Cir. 1997)) holds that "the disclosure of a process for obtaining cDNA from a particular organism and the description of the encoded protein fail to provide an adequate written description of the actual cDNA from that organism, despite the disclosure of a cDNA encoding that protein from another organism." First, it is necessary to clarify what *Eli Lilly* actually held. The Federal Circuit described the relationship of the written description requirement claims to genetic materials as follows:

> In claims to genetic material, however, a generic statement such as "vertebrate insulin cDNA" or "mammalian insulin cDNA," without more, is not an adequate written description of the genus because it does not distinguish the claimed genus from others, except by function. It does not specifically define any of the genes that fall within this definition. It does not define any structural features commonly possessed by members of the genus to distinguish them from others. One skilled in the art therefore cannot, as one can do with a fully described genus, visualize or recognize the identity of the members of the genus.

Id. at 1568. According to the court, the terms mammalian and vertebrate insulin cDNA identify a genus, but do not provide an adequate written description of the genus. The court then suggested that the description of a genus of cDNAs may be accomplished by defining the genetic

material by nucleotide sequence or by recitation of structural features common to the genus. *Id.* at 1569.

This is precisely what Appellants have done. They have defined the genetic material by reciting a specific sequence, SEQ ID NO:2, in the claims. The Appellants have not done what is impermissible under *Eli Lilly* - define the genetic material solely by function. In paragraphs 42, 44 and 45, the Office admits that such sequences have been disclosed, but ignores the fact that the claims have been limited by the recitation of a particular sequence and not just by reference to a function. Thus, the Office's arguments regarding *Eli Lilly* are misguided.

b. The Office failed to rebut the Appellants arguments based on the Written Description Guidelines

In their Appeal Brief at pages 14-15, the Appellants argued that the claims at issue to the claim analyzed in Example 14 of the USPTO's "Synopsis of Application of Written Description Guidelines". This example contains an analysis of claim that is highly similar to the claims at issue and which was found to be supported by an adequate written description as described in the Appeal Brief. The Appellants attached a copy of the relevant pages of the Written Description Guidelines to the Appeal Brief. However, the Examiner's Answer inexplicably misinterpreted this argument and thus failed to rebut it.

In paragraph 47 of the Examiner's Answer, the Office argued that it examined the specification and could not find an example 14 and that there is no disclosure of protein having 95% identity to SEQ ID NO:3. Thus, instead of referring to example 14 of the Written Description Guidelines, the Office attempted to find a non-existent example 14 in the specification. This mistake is inexplicable because the Appellants clearly argued in the Appeal

Brief and in their preceding Office action responses that the claims were analogous to example 14 of the Written Description Guidelines. Appellants are at a loss as to why this argument has not been properly considered. As it stands, this argument and the evidence supporting it stand unrebutted by the Office and thus the Appellants are entitled to removal of this ground of rejection.

Appellants reiterate that the claim of Example 14 recites a protein having SEQ ID NO:3 and variants thereof that are at least 95% identical to SEQ ID NO:3 and catalyze the reaction of A->B. The disclosure of Example 14 provides a single species (SEQ ID NO:3) that has actually been reduced to practice, and describes an assay for identifying the variants having the desired catalytic activity. The analysis considers (1) whether the members of genus vary substantially from each other; and (2) whether the disclosed species is representative of the members of the genus; in order to determine whether one of skill in the art would determine if the applicant was in possession of the necessary common attributes possessed by the members of the genus.

For Example 14, it was determined that the member species did not substantially vary since the variants have 95% identity or greater to the reference sequence, and also possess the catalytic activity. It was also determined that the disclosed species was representative since all members must have at least 95% structural identity to SEQ ID NO:3. The analysis determined that one of skill in the art would conclude that the applicant was in possession of the necessary common attributes possessed by the members of the genus, and therefore the disclosure in this Example meets the written description requirement. Appellants submit that the polypeptides encoded by the polynucleotides of claims 73, 74 and 80 can be analyzed in a similar manner to that provided in Example 14. First, the polypeptides encoded by the polynucleotides do not substantially vary as members of a genus since they all have at least 80% (or 90%) identity to a

recited portion of SEQ ID NO:2 and possess the same binding activity. Second, the polypeptide having SEQ ID NO:2 is a representative species of the genus since all polypeptides must have at least 80% (or 90%) identity to this sequence. Therefore, one of skill in the art would conclude that the Appellants were in possession of the necessary common attributes possessed by the members of the genus, and therefore the instant specification meets the written description requirement for these claims.

Claims 84 - 89 are directed to host cells, vectors, and methods utilizing the sequences of Claim 73 and thus the same reasoning that applies to Claim 73 applies to these Claims. Appellants note that the Office has not addressed these Claims separately thus believe that the additional elements of these Claims do not raise additional written description issues.

In light of the arguments set forth above, Appellant respectfully requests that the Office reconsider and withdraw the rejections of the claims on the basis of the 35 U.S.C. § 112, first paragraph, written description requirement.

3. The Claims Are Not Obvious

Claims 73 - 78 and 80 - 89 stand rejected under 35 U.S.C. § 103(a) as allegedly obvious over Valiante et al. (U.S. Pat. No. 5,688,690), Sambrook et al. (Molecular Cloning - A Laboratory Manual, 2nd Edition, Cold Spring Harbor, N.Y. 1989, pp. 2.43 - 2.84) and Porunellor et al. (J. Immunol. (1993) 151:5328 - 5337). Appellant respectfully maintains its argument that this rejection is in error because the Office failed to establish a *prima facie* case of obviousness with respect to any of the pending Claims. In particular, Appellant submits that 1) the Office is still confusing claims to nucleic acid sequences with claims to proteins; 2) the Examiner's Reply fails to rebut Appellants arguments regarding the controlling case of *In re Deuel*; and 3) the

Office has failed to rebut the Appellants arguments establishing that there was no motivation to combine the references.

a. The Claims are to nucleic acids, not proteins

In their Appeal Brief, the Appellants demonstrated that the Office has continually confused the claimed nucleic acid sequences with protein sequences. In particular, the Appellants pointed out the primary mistake made by the Office is highlighted in the Advisory Action where the Office stated that "it is the Appellants burden to approve [sic] that the claimed **polypeptide** and the p38 disclosed by Valainte et al. are isolated from the same source (human NK cells stimulated with IL2, IL12), have the same molecular weight (p38 Kd) and are recognized by the same monoclonal antibody (CI.7)." (Paper No. 12, p. 5, emphasis added).

Appellants must again emphasize that the present claims are to **nucleic acid sequences, not polypeptides**. The Examiner's Reply ignores this argument and indeed, repeats the mistake of confusing nucleotide sequences with protein sequences. This confusion is highlighted by the Office's statement in paragraph 62:

> A patent cannot issue to a genus of degenerate nucleic acid molecule where the protein polypeptide encoded thereby was known prior to the invention was filed. Because if a protein was known in the art, the structure is an inherent characteristic of protein. It does not add any patentable weight for an invention for just sequencing a know [sic] protein in the art.

In the first sentence quoted, the Office recognizes that the claims are to nucleic acid sequences that encode a polypeptide. However, the mistake made by the Office is apparent in the second quoted sentence, where the Office speaks about protein structure. Taking this reasoning at face value, the Office is arguing that proteins have inherent structural characteristics, and is confusing such protein characteristics with characteristics of the claimed nucleic acids. Appellants fail to

understand how inherent structural characteristics of a protein are relevant to an analysis of nucleic acids. As explained in more detail in the next section, the case law is clear that a known protein does not obviate claims to nucleic acid sequences encoding the protein. As to the last sentence, the Appellants are claiming nucleic acid sequences, so the statement that there is no patentable weight for sequencing a known protein is not relevant. The Appellants did not sequence the protein, they cloned the gene encoding the protein.

b. *In re Deuel* controls

In their Appeal Brief as well in their previous papers, the Appellants established that the Office's rejection of the claims at issue as obvious is factually identical to the facts of *In re Deuel,* 34 USPQ2d 1210 (Fed. Cir. 1995). The basis of the Appellants' arguments, which is presented in detail in the Appeal Brief, is that *In re Deuel* holds that claims to a nucleic acid sequence are not rendered obvious by the disclosure of protein that the nucleic acid encodes. Thus, it is submitted that the largely uncharacterized proteins disclosed in Valiante and Porunellor do not render the claimed nucleic acid sequences obvious. The Office has again failed to rebut this argument by establishing that the facts are different than has been argued or that some other case than *In re Deuel* controls.

In particular, the Office addressed, but did not rebut, Appellants In re Deuel argument in paragraph 60 of the Examiner's Reply:

> Appellants' argument has been respectfully considered; however, it is not found persuasive because claimed invention is drawn to polynucleotide of SEQ ID NO:1 isolated from cDNA library of human NK cells and human NK cells stimulated with IL-2, IL-12, IL-15, INF-g or anti-CD16 by using the commercial monoclonal Ci.7 (ATCC HB 117170) disclosed by Valiante et al. The cDNA of SEQ ID NO:1 encodes a 38 Kd protein NAIL of SEQ ID NO:2, which is highly expressed in human NK cells and monocytic cell line U937 followed by CD8+ cells. NAIL binds to CD48 and the interaction of NAIL with CD48 enhances the activation of NK cells.

This argument (and the additional arguments in paragraphs 61 and 62) does nothing to address the facts or legal holding of *In re Deuel*. Indeed, the Office has failed to explain why *In re Deuel* is not controlling. Thus, Appellants are entitled to removal of the obviousness rejection.

An alternative way to analyze this issue is to consider that cited references do not provide an adequate written description of the claimed nucleic acid sequences. A recent Federal Circuit case, *Noelle v. Lederman,* 2004 U.S. App. LEXIS 774 (Fed. Cir. 2004), holds that claims to a human antibody are not sufficiently disclosed for purpose of the written description requirement where the specification did not describe structural elements of the human antibody or antigen. Id. at *15-16. The Federal Circuit quoted *Fiers v. Revel*, 984 F.2d 1164 (Fed. Cir. 1993) in holding that claims to DNA cannot be supported by a "mere wish or plan for obtaining the claimed chemical invention." As stated by the Federal Circuit:

> Therefore, this court has held that statements in the specification describing the functional characteristics of a DNA molecule or methods of its isolation do not adequately describe a particular claimed DNA sequence. Instead "an adequate written description of DNA requires more than a mere statement that it is part of the invention and reference to a potential method for isolating it; what is required is a description of the DNA itself." *Id. at 1566-67* (quoting *Fiers, 984 F.2d at 1171*).

Id. at *15 -*16 (quoting *Regents of the University of California v. Eli Lilly and Co.*, 119 F.3d 1559, 1566-67 (Fed. Cir. 1997) and *Fiers v. Revel*, 984 F.2d 1164, 1171 (Fed. Cir. 1993)).

This case is analogous to the present application where the Office argues that references that provide a "mere wish or plan" for obtaining the claimed chemical invention make the chemical invention obvious. This cannot be the case because the cited references do not provide any disclosure of the chemical structure of the claimed invention and thus cannot make the claimed invention obvious. This established legal principle is in direct contrast to the Office's reasoning advanced in paragraph 66 - 69 that Valiante and Sambrook teach how to clone the

claimed sequences. Appellants respectfully note that the Valiante example referred to in paragraph 68 is a prophetic example; the NAIL polynucleotide sequence apparently remained uncloned until Appellants efforts 4 years after the filing date of Valiante. Thus, the references relied on by the Office are inadequate to obviate the claimed invention because they provided nothing more than a "mere wish or plan" for obtaining the claimed invention.

c. The Office has not established a motivation to combine

The Office has similarly failed to rebut the arguments concerning motivation to combine in the Appellants' Appeal Brief. In paragraph 64 of the Examiner's Answer, the Office states that:

> It must be recognized that any judgment on obviousness is in a sense necessarily a reconstruction based on hindsight reasoning. But so long as it takes into account only knowledge which was within the level of ordinary skill at the time the claimed invention was made, and does not include knowledge gleaned only from the Appellants' disclosure, such a reconstruction is proper. See *In re McLaughlin*, 443 F.2d 1392, 170 USPQ 209 (CCPA 1971).

Applicant respectfully submits that the Office's reliance on *McLaughlin* is misguided and that hindsight reconstruction is legally impermissible. To the extent that this 1971 C.C.P.A. case appears to condone hindsight reconstruction when providing a motivation to combine references, the Federal Circuit has *sub silentio* overruled this proposition, and has emphatically stated that hindsight reconstruction is not proper (as detailed below).

The Federal Circuit has repeatedly warned against using hindsight reconstruction as a test of obviousness. A few examples of such cases include: *In re Fine*, 837 F.2d 1071 (Fed. Cir. 1988) ("One cannot use hindsight reconstruction to pick and choose among isolated disclosures in the prior art to deprecate the claimed invention"); *Gillette Co. v. S.C. Johnson & Son, Inc.*, 919 F.2d 720 (Fed. Cir. 1990) (The inappropriateness of hindsight as a test of obviousness was, in

point of fact, discovered, and articulated lucidly, over three centuries ago, by Milton, who, in Paradise Lost Part IV, L. 478-501, stated "The invention all admired, and each how he To be the inventor missed; so easy it seemed, Once found, which yet unfound would have thought, Impossible!"); *Heidelberger Druckmaschinen AG v. Hantscho Commercial Products, Inc.*, 21 F.3d 1068 (Fed. Cir. 1993) ("The motivation to combine references can not come from the invention itself"); *Sensonics, Inc. v. Aerosonic Corp.*, 81 F.3d 1566 (Fed. Cir. 1996) ("To draw on hindsight knowledge of the patented invention, when the prior art does not contain or suggest that knowledge, is to use the invention as a template for its own reconstruction-an illogical and inappropriate process by which to determine patentability"); *W.L. Gore & Assocs., Inc. v. Garlock Inc.*, 721 F.2d 1540 (Fed. Cir. 1983) ("To imbue one of ordinary skill in the art with the knowledge of the invention in suit, when no prior art reference or references of record convey or suggest that knowledge, is to fall victim to the insidious effect of hindsight syndrome wherein that which only the inventor taught is used against its teacher . . .").

The Office's combination is clearly based on "knowledge gleaned from the Appellants' disclosure." The Office has made a conceptual leap that is unsupported by the references themselves. This unsupported leap is pure hindsight reconstruction as evidenced by the Office's statement in paragraph 67 that the "person of ordinary skill in the art would have been motivated by the cited references to search this **conserved** cell surface signaling molecule because the non-MHC restricted NK cell mediated killing is very important in the in the field as taught by Valiante et al and Porunelloor et al. " (Emphasis added). The clear mistake of the Office is the conclusion that the sequences of the two references are conserved where there is absolutely no disclosure in the references themselves that the sequences are related or conserved. The fact that the sequences are conserved could only have come from the Appellants disclosure.

Moreover, Porunelloor actually teaches away from the expression of a similar, conserved, gene in human NK cells. At page 5330, column 1, Porunelloor discloses that Northern blots conducted using human RNA indicate that a 2B4 homolog is not expressed in humans:

> Genomic Southern blots identified a human homologue of the 2B4 gene. However, RNA blot analysis of total RNA isolated from human NK cells suggests that 2B4 gene is not expressed in humans.

The Federal Circuit has made it clear that there can be no suggestion to combine or modify "if the reference teaches away from its combination with another source." *Tec Air, Inc. v. Denso Manufacturing Michigan, Inc.*, 192 F.3d 1353 (Fed. Cir. 1999). Because Porunelloor teaches that a human homolog is not expressed, a person of skill in the art would not be motivated to combine Porunelloor with Valiante, which is directed to human protein. In other words, a person of ordinary skill in the art would not believe that a 2B4 homolog encodes the human p38 protein of Valiante because Porunelloor could not detect expression of a 2B4 homolog in human cells.

Finally, in paragraph 68, the Office states that Valiante et al. teach "working examples" of the cloning of p38. However, as previously described, these examples are prophetic and the Office has provided no evidence that such methods are more than "mere wish or plan" for cloning the claimed sequences. Indeed, the known evidence is to the contrary since it appears that Valiante was not successful in cloning the claimed sequences in the four year period between the filing date of Valiante et al. and the present application.

For the foregoing reasons, Appellants reiterate their argument that the Office has failed to provide a motivation to combine the cited references.

E. Conclusion

For the foregoing reasons, it is submitted that the Office's rejection of Claims 73 - 78 and 80 - 89 was erroneous, and reversal of the rejection is respectfully requested. Appellant requests either that the Board render a decision as to the allowability of the claims, or alternatively, that the application be remanded for reconsideration by the Office.

Dated: 3-8-04

J. Mitchell Jones
Registration No. 44,174

MEDLEN & CARROLL, LLP
101 Howard Street, Suite 350
San Francisco, California 94105

APPENDIX A

PENDING CLAIMS

The following is a list of the pending Claims.

73. An isolated nucleic acid molecule comprising a polynucleotide encoding a polypeptide at least 80% identical to amino acids 22-221 of SEQ ID NO:2, wherein the polypeptide binds CD48.

74. An isolated nucleic acid molecule of claim 73, wherein the polypeptide acid sequence is at least 90% identical to amino acids 22-221 of SEQ ID NO:2, wherein the polypeptide binds CD48.

75. The isolated nucleic acid molecule of claim 73, wherein the polypeptide comprises amino acids 22-221 of SEQ. ID NO:2.

76. The isolated nucleic acid molecule of claim 73, wherein the polypeptide comprises amino acids 1-221 of SEQ ID NO:2.

77. The isolated nucleic acid molecule of claim 73, wherein the polypeptide comprises amino acids 19-221 of SEQ ID NO:2.

78. The isolated nucleic acid molecule of claim 73, wherein the polypeptide comprises amino acids 19-224 of SEQ ID NO:2.

80. An isolated nucleic acid molecule comprising a polynucleotide at least 80% identical to SEQ ID NO:1.

81. The isolated nucleic acid molecule of claim 73, wherein the polypeptide comprises SEQ ID NO:6.

82. The isolated nucleic acid molecule of claim 73, wherein the polypeptide comprises SEQ ID NO:7.

83. The isolated nucleic acid molecule of claim 73, wherein the polypeptide comprises SEQ ID NO:8.

84. A recombinant vector comprising the nucleic acid molecule of any one of claims 73 through 83.

85. A host cell transfected or transduced with the vector of claim 84.

86. A method for the production of NK cell Activation Ligand (NAIL) polypeptide comprising culturing a host cell that has been genetically engineered to express the nucleic acid of claim 73 under conditions promoting expression of the polypeptide.

87. The method of claim 86, further comprising recovering the polypeptide.

88. The method of claim 87, wherein the host cell is a mammalian cell.

89. The method of claim 88, wherein the host cell is a CV-1/EBNA cell.

Improved Green Fluorescent Protein by Molecular Evolution Using DNA Shuffling

Andreas Crameri, Erik A. Whitehorn, Emily Tate, and Willem P. C. Stemmer*

*Affymax Research Institute, 4001 Miranda Ave., Palo Alto, CA 94304. *Corresponding author (e-mail: pim_stemmer@qmgates.affymax.com).*

Received 10 November 1995; accepted 27 December 1995.

Green fluorescent protein (GFP) has rapidly become a widely used reporter of gene regulation. However, for many organisms, particularly eukaryotes, a stronger whole cell fluorescence signal is desirable. We constructed a synthetic GFP gene with improved codon usage and performed recursive cycles of DNA shuffling followed by screening for the brightest *E. coli* colonies. A visual screen using UV light, rather than FACS selection, was used to avoid red-shifting the excitation maximum. After 3 cycles of DNA shuffling, a mutant was obtained with a whole cell fluorescence signal that was 45-fold greater than a standard, the commercially available Clontech plasmid pGFP. The expression level in *E. coli* was unaltered at about 75% of total protein. The emission and excitation maxima were also unchanged. Whereas in *E. coli* most of the wildtype GFP ends up in inclusion bodies, unable to activate its chromophore, most of the mutant protein is soluble and active. Three amino acid mutations appear to guide the mutant protein into the native folding pathway rather than toward aggregation. Expressed in Chinese Hamster Ovary (CHO) cells, this shuffled GFP mutant showed a 42-fold improvement over wildtype GFP sequence, and is easily detected with UV light in a wide range of assays. The results demonstrate how molecular evolution can solve a complex practical problem without needing to first identify which process is limiting. DNA shuffling can be combined with screening of a moderate number of mutants. We envision that the combination of DNA shuffling and high throughput screening will be a powerful tool for the optimization of many commercially important enzymes for which selections do not exist.

Keywords: sexual PCR, molecular libraries, protein expression

The green fluorescent protein (GFP) of the jellyfish *Aequorea victoria* is a useful reporter for gene expression and regulation[1–3]. GFP has been expressed in a wide variety of microbial, plant, insect and mammalian cells[4–9]. However, for most assays a significant increase in the fluorescence signal from, particularly eukaryotic, cells expressing GFP would be very useful. The 4- to 6-fold improvement in excitation reported for a mutant GFP (S65T) is a step in the right direction, but even stronger signal is desirable for most applications[10,11]. GFP mutants with a red-shifted excitation spectrum[12,13], although also useful, require detection with special filter sets, such as FACS or fluorescence microscopy.

Our goal was to improve the whole cell fluorescence of GFP for general use as a reporter in prokaryotic and eukaryotic cells. The improvement of GFP by rational design appeared difficult because the quantum yield of GFP is already 0.7 to 0.8[14] and the expression level of GFP in our *E. coli* construct was already about 75% of the total protein. We first tried to improve GFP by synthesizing a GFP gene with optimized codon usage. We then attempted to further improve GFP using an evolutionary approach, consisting of recursive cycles of DNA shuffling[15–17] of the GFP gene, combined with visual selection of the brightest clones.

DNA shuffling is a technique for *in vitro* recombination of pools of homologous genes. The pool of genes is fragmented into random size pieces, and the PCR reassembly of full-length genes from the fragments via self-priming yields crossovers due to PCR template switching[15]. Coupled with selection or screening, this homologous recombination process is the most efficient known process for combining positive mutations and simultaneously removing negative mutations from the sequence pool[16,17]. UV light was used for the selection, in order to prevent red-shifting the excitation maximum which makes detection by the naked eye difficult. Since shuffling is technically simple in bacteria, we opted to optimize the whole cell fluorescence signal in *E. coli*, and then assay the performance of the best GFP mutants in eukaryotic cells.

Results and Discussion

While the wild type and cycle 1–3 GFP constructs differ only in a few mutations in the sequence of their GFP gene, the Clontech pGFP construct differs from all other constructs in the sequence of its GFP gene as well as in the expression vector. Improvements in the whole cell fluorescence over the Clontech pGFP construct thus cannot be attributed solely to the sequence of the GFP protein. However, comparison to the Clontech construct is relevant because it is commercially available and widely used, and it has become therefore the *de facto* standard to which newer constructs should be compared. The vector pGFP was originally described by Chalfie et al.[1] as TU#60, and encodes GFP with wild-type codons utilized in *A. victoria*.

E. coli expressing our synthetic GFP construct ('wt') with altered codon usage yielded a nearly 3-fold greater whole cell fluorescence signal than cells expressing the Clontech construct (Figs. 1A & B). The comparison was performed after complete induction and at equal cell densities.

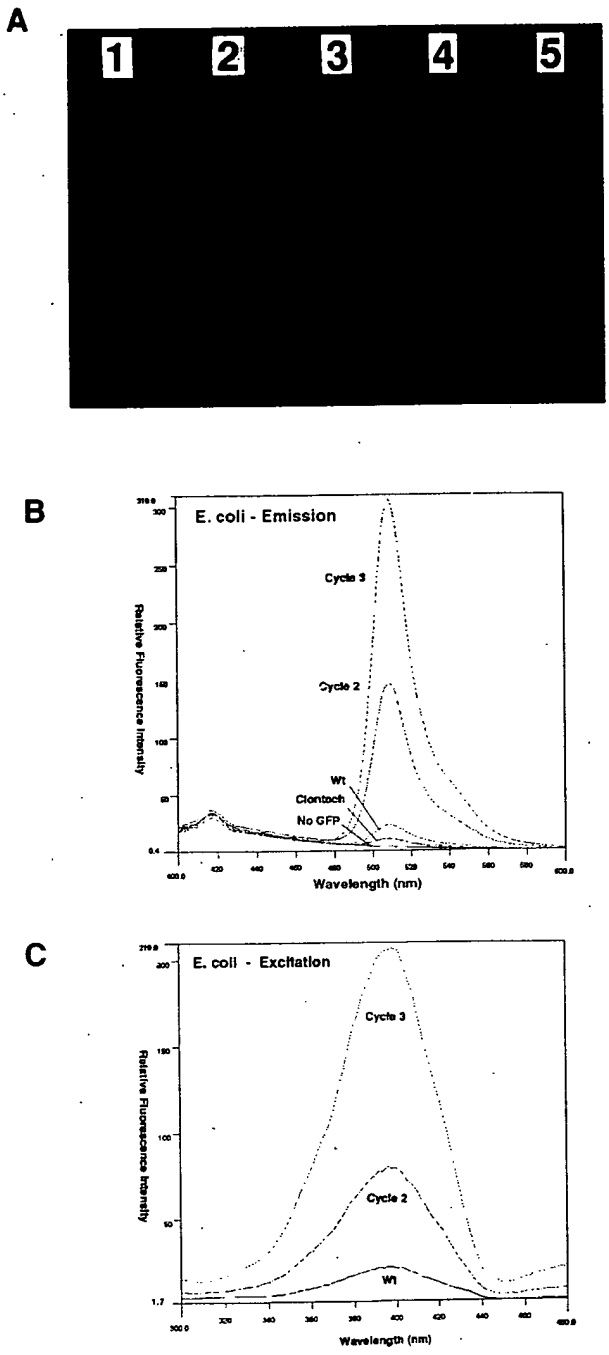

Figure 1. Comparison of the fluorescence of different GFP constructs in whole *E. coli* cells. Compared are the Clontech GFP construct with the amino acid sequence reported by Prasher et al.', our wildtype construct ('wt') with an amino acid sequence correction, an alanine insertion after the fMet as well as improved codon usage, and the mutants obtained after 2 or 3 cycles of DNA shuffling and selection ('cycle 2', 'cycle 3'). While the 'wt', 'cycle 2' and 'cycle 3' GFP mutants use identical vectors, the Clontech vector is quite different. For example, the 'Clontech' construct is induced with IPTG, whereas the other three constructs are induced with arabinose. All samples were assayed at equal OD. (A) Photograph of *E. coli* cell suspensions at equal densities over a 365 nm UV light box. No filters were used. (1) no GFP, (2) Clontech GFP, (3) wt GFP, (4) cycle 2 mutant GFP, (5) cycle 3 mutant GFP. (B) Fluorescence spectra show that the whole cell fluorescence signal from the 'wt' construct is 2.8-fold greater than from the Clontech construct. The signal of the 'cycle 2' mutant is 8-fold greater than the 'wt' construct, and 22-fold over the Clontech construct. The signal of the 'cycle 3' mutant is 16-fold greater than the 'wt' construct, and 45-fold over the 'Clontech' construct. (C) Comparison of excitation spectra of GFP constructs in *E. coli*.

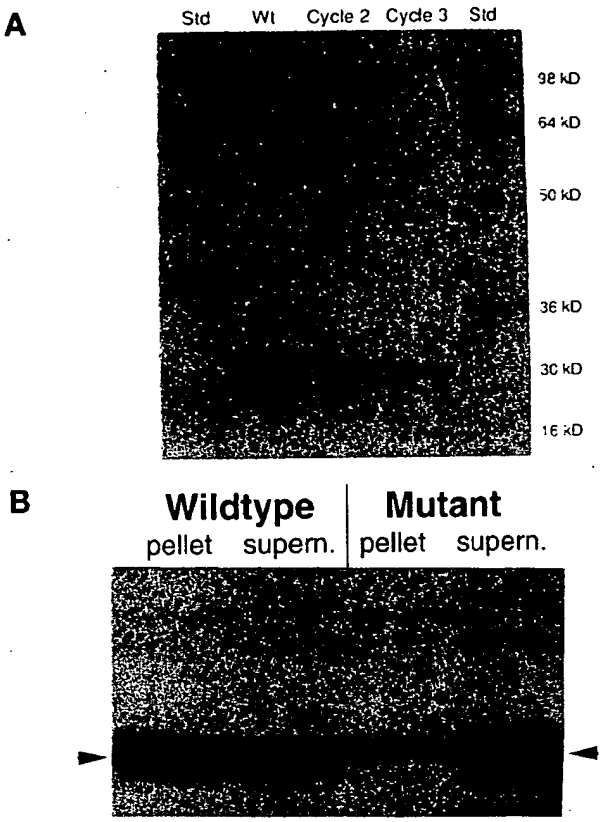

Figure 2. (A) A 12% Tris-glycine SDS-PAGE analysis of equal amounts whole *E. coli* cells expressing the wildtype, the cycle 2 mutant or th cycle 3 mutant of GFP. Stained with Coomassie Blue. (B) A 12% Tris glycine SDS-PAGE analysis of equal amounts of *E. coli* fractions. Lan 1: Pellet of lysed cells expressing wt GFP; Lane 2: Supernatant of lyse cells expressing wt GFP (most of the wt GFP is in inclusion bodies); Lan 3: Pellet of lysed cells expressing cycle 3 mutant GFP; Lane Supernatant of lysed cells expressing cycle 3 mutant GFP (most of th wt GFP is soluble).

DNA shuffling. The fluorescence signal of the synthetic 'wt' GF was further improved using DNA shuffling to construct a muta library as described in the Experimental Protocol, followed by pla ing and selecting the brightest colonies. After the second cycle shuffling and selection, a mutant ('cycle 2') was obtained that w about 8-fold improved over 'wt', and 23-fold over the Clontech co struct. After the third cycle a mutant ('cycle 3') was obtained whi was 16- to 18-fold improved over the 'wt' construct, and 45-fold ov the Clontech construct (Figs. 1A & B). The peak wavelengths of t excitation and emission spectra of the mutants were identical to th of the 'wt' construct (Figs. 1B & C).

SDS-PAGE analysis of whole cells showed that the total level of t GFP protein expressed in all three constructs was unchanged, a surprisingly high 75% of total protein (Fig. 2A). Fractionation of t cells by sonication and centrifugation showed that the 'wt' constru contained mostly inactive GFP in the form of inclusion bodi whereas the 'cycle 3' mutant GFP remained mostly soluble and w able to activate its chromophore.

The mutant genes were sequenced and unexpectedly, the 'cycl mutant was found to contain more mutations than the 'cycle mutant (Fig. 3). The 'cycle 3' contained 3 protein mutations an silent mutations relative to the 'wt' construct. Mutations F10 M154T, and V164A involve the replacement of hydrophobic resid with more hydrophilic residues[18]. In jellyfish, GFP is known to b to a protein called Aequorin, most likely involving a hydroph contact area. In the absence of Aequorin, this hydrophobic surface

may cause aggregation and prevent autocatalytic activation of the chromophore. The three hydrophylic mutations that were obtained may counteract the hydrophobic site, resulting in reduced aggregation and increased chromophore activation.

Pulse chase experiments with whole bacteria grown at 37°C showed that the $T_{1/2}$ for fluorophore formation was about 95 minutes for both the 'wt' and the 'cycle 3' mutant GFP (Fig. 4). The $T_{1/2}$ of 4hours reported previously was obtained at 22°C and after anaerobic growth[10].

Growth rate enhancement. In addition to increases in the fluorescence signal, we found that *E. coli* colonies expressing the 'cycle 2' and 'cycle 3' GFP mutants grew 2- to 3-fold faster than colonies expressing the 'wt' construct (data not shown). Such an increase in growth rate, presumably due to reduced toxicity of the overexpressed gene, is an additional benefit of using a screen or selection which can broadly and simultaneously improve many factors.

CHO cells. After being selected in bacteria, the 'cycle 3' mutant GFP was transferred into the eukaryotic Alpha+ vector and expressed in chinese hamster ovary cells (CHO) (Fig. 5). While we hoped that the improvement of the folding of GFP would be transferable to mammalian cells, we were nonetheless surprised to find that CHO cell clones expressing the 'cycle 3' mutant GFP, which in *E. coli* gave a 16-fold stronger signal than the 'wt' construct , yielded a 42-fold greater whole cell fluorescence signal than clones expressing the 'wt' construct (Fig. 6A). FACS analysis confirmed that the average fluorescence signal of CHO cell clones expressing 'cycle 3' was 46-fold greater than cells expressing the 'wt' construct (Fig. 6B). The addition of 2 mM sodium butyrate increased the fluorescence signal of both constructs about 2- to 8-fold. While we focused on comparing the best CHO cell clones obtained with each GFP mutant, a similar improvement was observed in the cell population before clone selection (data not shown). Transient expression of the different GFP mutants also showed that the 'cycle 3' mutant was clearly detectable one and two days after electroporation, whereas the 'wt' GFP construct was not (data not shown).

Screening versus selection. These results were obtained by visually screening approximately 10,000 *E. coli* colonies per cycle and picking the brightest 40 colonies to use in the next cycle. An important question is whether significant improvements in protein function can routinely be obtained with such low numbers. If so, the combination of DNA shuffling with high-throughput screening may become a powerful process for the optimization of the large number of commercially important enzymes for which selections are not feasible.

Comparison to other mutants. A different GFP improvement was recently obtained in *E. coli* by Cormack et al. after high level synthetic mutagenesis of a 20 amino acid window around the chromophore combined with a FACS selection, which yielded mutants with a red-shifted excitation spectrum[12].

We compared Cormack's brightest mutant (M2) with our pBAD-GFP cycle 3 mutant (C3) in *E. coli*. Because the expression plasmids and induction conditions differ significantly, we also cloned a 65 bp *NcoI-NdeI* restriction fragment containing their GFP mutations into our wildtype GFP plasmid (WT), resulting in pBAD-GFP WT/M2 (WT/M2). Cells were grown under identical conditions and induced under conditions optimal for each construct, as previously described. Using excitation at 385 nm, we measured the whole cell fluorescence at 510 nm at equal cell density (OD_{600}).

In *E. coli*, mutant C3 yielded a 9.2-fold stronger fluorescence signal than M2, and the WT/M2 construct yielded a 3.1-fold stronger fluorescence signal than M2. In each of these comparisons the plasmid and GFP sequences differ in many respects and no additional conclusions can be drawn. From SDS-PAGE gels the expression level of GFP appears to be nearly 2-fold higher for C3 compared to M1–3.

In a direct comparison of the three amino acid mutations

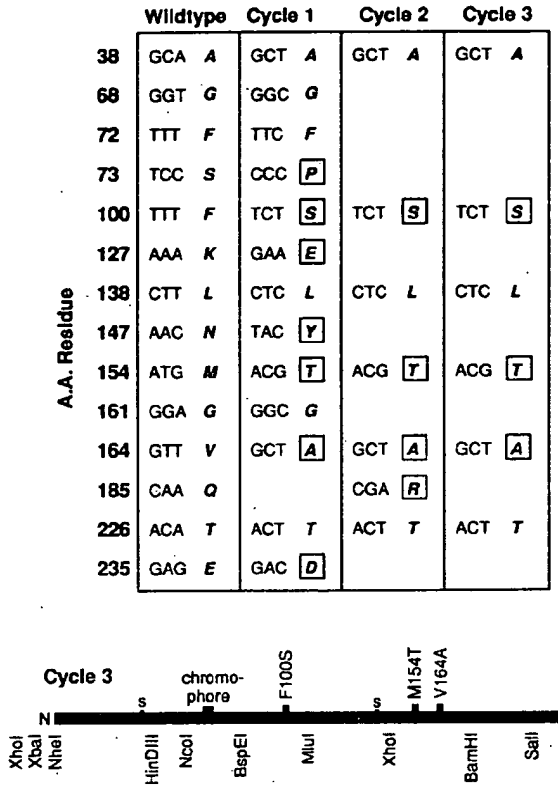

Figure 3. Mutation analysis of the cycle 2 and cycle 3 mutants versus wildtype GFP.

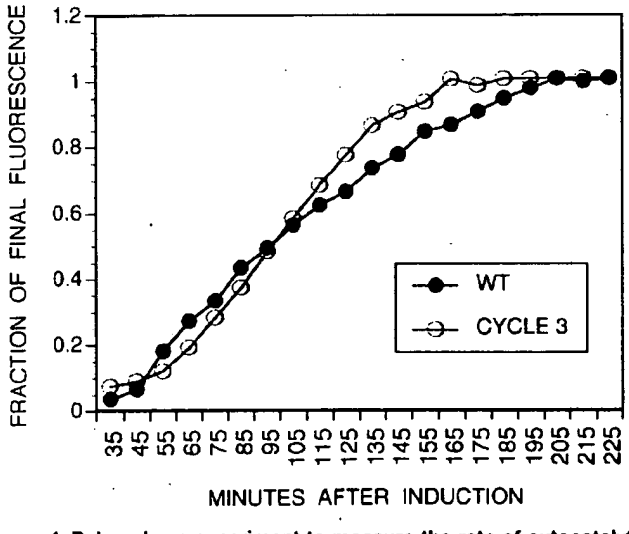

Figure 4. Pulse-chase experiment to measure the rate of autocatalytic activation of the GFP chromophore. *E. coli* cells expressing the 'wt' or the 'cycle 3' mutant of GFP were grown at 37°C, repressed with 2% glucose, to an OD_{600} of 0.35. The cells were centrifuged and induced by transfer to medium containing 0.2% arabinose. After 30 minutes of induction, protein synthesis was stopped by addition of 100 µg/ml chloramphenicol. The timescale was calculated from the middle of the induction period.

obtained throughout the protein by shuffling with the three mutations obtained by cassette mutagenesis near the chromophore, C3 yielded a 2.9-fold improvement over the WT/M2 construct.

While our data suggests that pBAD-GFP cycle 3 is the best for UV detection, when excitation at 488 nm is used, the M2 mutant in *E. coli* is 4.1-fold brighter than C3 and should therefore be better for FACS selection. The comparison can only be done for *E. coli* because

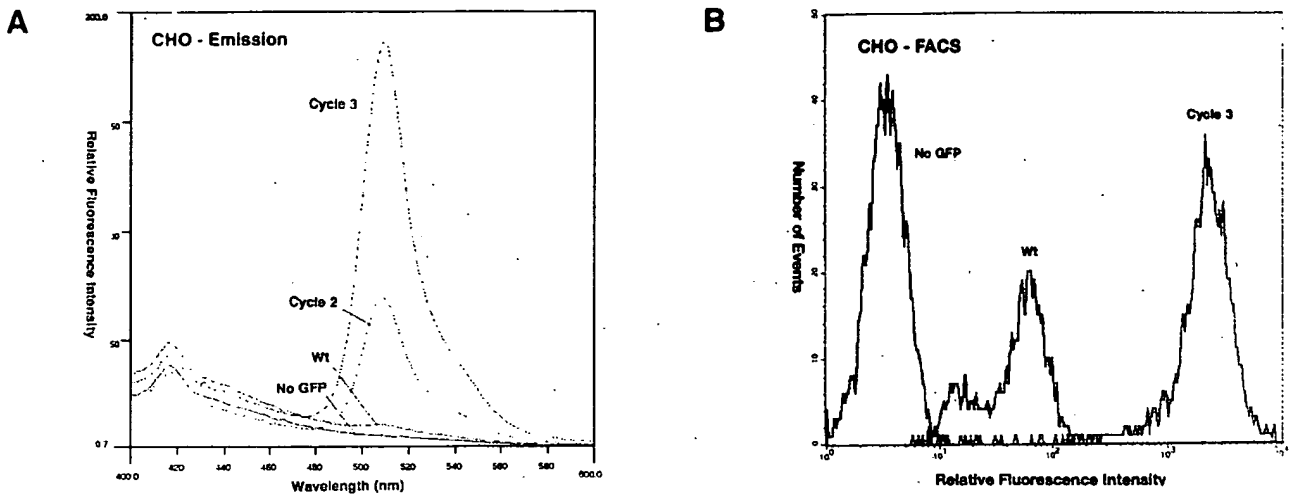

Figure 6. (A) Fluorescence spectroscopy of clones of CHO cells expressing different GFP mutants. The signal of the cells expressing the 'cyc GFP mutant is 16-fold greater than the 'wt' construct. The signal of cells expressing the 'cycle 3' GFP mutant is 42-fold greater than the 'wt' struct. (B) FACS analysis of clones of CHO cells expressing different GFP mutants. The average fluorescence intensity of cells expressing 'cycle 2' GFP mutant is 26-fold greater than the 'wt' construct. The average fluorescence intensity of the 'cycle 3' GFP mutant is 46-fold gre than the 'wt' construct.

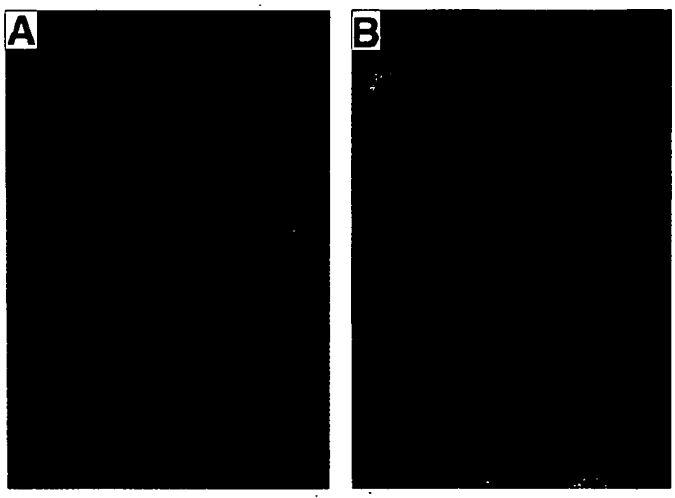

Figure 5. (A) CHO cells expressing 'wt' GFP. (B) CHO cells expressing 'cycle 3' mutant GFP. Living cells grown on coverslips were photographed using an oil immersion objective.

Cormack et al. did not report the performance of their mutants in mammalian cells[12].

As a practical matter, for most routine applications red-shifted mutants are more difficult to detect. Such mutants need to be excited with visible light, and the much weaker fluorescence emission cannot be seen with the naked eye due to this excess of visible light. Therefore special optical filters sets, such as used in FACS or fluorescence microscopy, are required to differentiate the 510 nm emitted light from the 488 nm light used for excitation.

Experimental Protocol

GFP gene construction. We obtained the commercially available pGFP plasmid from Clontech. This gene has the GFP sequence reported by Chalfie et al.[3], and contains a Q80R mutation which occurred as a PCR error as well as 24 extra amino acids from the N-terminus of LacZ (Clontech, personal communication). Expression of GFP from the 'Clontech' construct is inducible with IPTG. We synthesized a GFP gene with the published amino acid sequence but without the 24 residue N-terminal addition of LacZ, but including the Q80R mutation as well as an alanine residue insertion after the fMet. This GFP sequence is referred to as our wildtype ('wt'). The gene was synthesized 14 oligonucleotides ranging from 54 to 85 bases which were assembled as s pairs by PCR extension. These segments were digested with restric enzymes and cloned separately into the vector Alpha+[19] and sequenced. T segments were then ligated into the eukaryotic expression vector Alpha form the full-length GFP construct, Alpha+GFP (Fig. 7). Arginine codons are known to be poorly translated in *E. coli* were replaced at amino acid p tions 73 (CGT), 80 (CGG), 96 (CGC) and 122 (CGT). A number of other s mutations were engineered into the sequence to create the restriction sites in the assembly of the gene. These were S2 (AGT to AGC; to create an *Nhe*I K41 (AAA to AAG; HinDIII), Y74 (TAC to TAT) and P75 (CCA to C BspEI), T108 (AGA to AGG; MluI), L141 (CTC to TTG) and E142 (GA GAG; *Xho*I), S175 (TCC to AGC; *Bam*HI) and S202 (TCG to TCC; *Sal*I) E235 (GAA to GAG) and L236 (CTA to CTC; *Sac*I). The 5′ and 3′ untrans ends of the gene contained *Xba*I and *Eco*RI sites, respectively. The sequen the gene was confirmed by sequencing. The *Xba*I-*Eco*RI fragmen Alpha+GFP, containing the whole GFP gene, was subcloned into the prok otic expression vector pBAD18[20], resulting in the bacterial expression ve pBAD18-GFP (Fig. 7). In this vector GFP gene expression is under the co of the arabinose promoter/repressor (araBAD), which is inducible arabinose (0.2%).

Gene shuffling and selection. An approximately 1 kb DNA fragment taining the whole GFP gene was obtained from the pBAD-GFP tor (Fig. 1) by PCR with primers TAGCGGATCCTACCTGACGC (nea *Nhe*I site) and GAAAATCTTCTCTCATCCG (near the *Eco*RI and purified by Wizard PCR prep (Promega, Madison, WI). This PCR pro was digested into random fragments with DNase I (Sigma) and 50 to 30 fragments were purified from 2% low melting point agarose gels. The pur fragments were resuspended at 10 to 30 ng/μl in a PCR mixture containin mM each dNTP, 2.2 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-HCl pH 9.0, Triton-X-100 with Taq DNA polymerase (Promega, Madison, WI) and as bled (without primers) using a PCR program of 35 cycles of 94°C 30s, 30s, 72°C 30s, as described previously[16]. The product of this reaction diluted 40X into new PCR mix, and the full length product was amplified the same two primers in a PCR of 25 cycles of 94°C 30s, 50°C 30s, 72°C 30 lowed by 72°C for 10 min. After digestion of the reassembled product *Nhe*I and *Eco*RI, this library of point-mutated and *in vitro* recombined genes was cloned back into the pBAD vector[20]. The ligated DNA was ele porated into *E. coli* TG1 (Pharmacia) which were plated on LB plates with μg/ml ampicillin and 0.2% arabinose to induce GFP expression from the binose promoter.

Mutant selection. Over a standard UV light box (365 nm) the 40 brig colonies were selected and pooled. These colonies were used as the templa a PCR reaction to obtain a pool of GFP genes. Cycles 2 and 3 were perfo

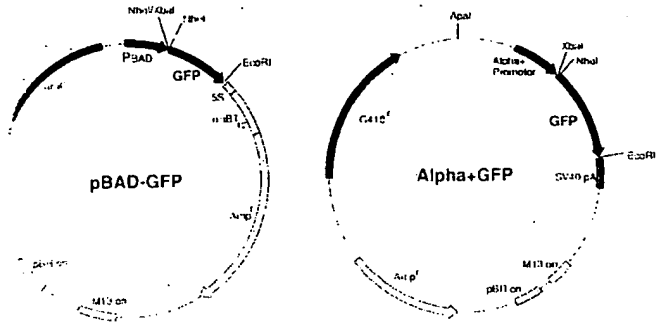

Figure 7. The prokaryotic GFP expression vector pBAD-GFP (5,371 bp) was derived from pBAD18[20]. The eukaryotic GFP expression vector Alpha+GFP (7,591 bp) was derived from the vector Alpha+[19].

identical to cycle 1. The best mutant from cycle 3 was identified by fluorescence spectrometry of colonies grown in microtiter plates.

Mutant characterization in *E. coli*. DNA sequencing was performed on an Applied Biosystems 391 DNA sequencer. Fractionation of *E. coli* cells into inclusion bodies and soluble protein was performed by sonication for 1 min followed by centrifugation at 10,000 g.

CHO cell expression of GFP. The wildtype and the cycle 2 and 3 mutant versions of the GFP gene were transferred into the eukaryotic expression vector Alpha+[19] as an *Eco*RI-*Xba*I fragment (Fig. 7). The plasmids were transfected into CHO cells by electroporation of 8×10^6 cells in 0.8 ml with 40 μg of plasmid at 400V and 250 μF. Transformants were selected using 1 mg/ml G418 for 10 to 12 days[18]. Cells were treated with 2 mM sodium butyrate for 36 to 48 h before their fluorescence was observed.

FACS analysis. FACS analysis was carried out on a Becton Dickinson FACSTAR Plus using an Argon ion laser tuned to 488 nm. Fluorescence was observed with a 535/30 nm bandpass filter.

Fluorescence spectroscopy. Whole cell fluorescence spectra were obtained with a Perkin-Elmer LS50B luminescence spectrophotometer. *E. coli* cultures were measured at equal OD_{600}, and mammalian cells at equal cell numbers.

1. Prasher, D. C., Eckenrode, V. K., Ward, W. W., Prendergast, F. G. and Cormier, M. J. 1992. *Gene* **111**:229–233.
2. Prasher, D. C. 1995. Using GFP to see the light. *Trends in Genetics* **11**:320–323.
3. Chalfie, M. Tu, Y., Euskirchen, G., Ward, W. W. and Prasher, D. C. 1994. Green fluorescent protein as a marker for gene expression. *Science* **263**:802–805.
4. Inouye, S. and Tsuji, F. I. 1994. *FEBS Lett.* **341**:277–280.
5. Wang, S. and Hazelrigg, T. 1994. Implications for *bcd* mRNA localization from spatial distribution of *exu* protein in Drosophila oogenesis. *Nature* **369**:400–403.
6. Brand, A. 1995. GFP in *Drosophila*. *Trends in Genetics* **11**:324–325.
7. Pines, J. 1995. GFP in mammalian cells. *Trends in Genetics* **11**:326–327.
8. Hodgkinson, S. 1995. GFP in *Dictyostelium*. *Trends in Genetics* **11**:327–328.
9. Haseloff, J. and Amos, B. 1995. GFP in plants. *Trends in Genetics* **11**:328–329.
10. Heim, R., Prasher, D. C. and Tsien, R. Y. 1994. Wavelength mutations and post-translational autoxidation of green fluorescent protein. *Proc. Natl. Acad. Sci. USA* **91**:12501–12504.
11. Heim, R., Cubitt, A. B. and Tsien, R. Y. 1995. Improved green fluorescence. *Nature* **373**:663–664.
12. Cormack, B. P., Valdivia, R. H. and Falkow, S. 1996. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene. In press.*
13. Delagrave, S., Hawtin, R. E., Silva, C. M., Yang, M. M. and Youvan, D. C. 1995. Red-shifted excitation mutants of the green fluorescent protein. *Bio/Technology* **13**:151–154.
14. Ward, W. W., Cody, C. W., Hart, R. C. and Cormier, M. J. 1982. Spectrophotometric identity of the energy transfer chromophores in Renilla and Aequorea green fluorescent proteins. *Photochem. Photobiol.* **35**:803–808.
15. Stemmer, W. P. C. 1994. DNA shuffling by random fragmentation and reassembly: *In vitro* recombination for molecular evolution. *Proc. Natl. Acad. Sci. USA* **91**:10747–10751.
16. Stemmer, W. P. C. 1994. Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370**:389–391.
17. Stemmer, W. P. C. 1995. Searching sequence space. *Bio/Technology* **13**:549–553.
18. Kyte, J. and Doolittle, R. F. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**:105–132.
19. Whitehorn, E., Tate, E., Yanofsky, S. D., Kochersperger, L., Davis, A., Mortensen, R. B., Yonkovich, S., Bell, K., Dower, W. J. and Barrett, R. W. 1995. A generic method for expression and use of 'tagged' soluble versions of cell surface receptors. *Bio/Technology* **13**:1215–1219.
20. Guzman, L.-M., Belin, D., Carson, M. J. and Beckwith, J. 1995. Tight regulation, modulation, and high-level expression by vectors containing the arabinose P_{BAD} promoter. *J. Bacteriol.* **177**:4121–4130.

J. Mol. Biol. (1997) 272, 336–347

JMB

Strategies for the *in vitro* Evolution of Protein Function: Enzyme Evolution by Random Recombination of Improved Sequences

Jeffrey C. Moore, Hua-Ming Jin, Olga Kuchner and Frances H. Arnold*

Division of Chemistry and Chemical Engineering, Mail Code 210-41, California Institute of Technology Pasadena, CA 91125, USA

**Corresponding author*

Sets of genes improved by directed evolution can be recombined *in vitro* to produce further improvements in protein function. Recombination is particularly useful when improved sequences are available; costs of generating such sequences, however, must be weighed against the costs of further evolution by sequential random mutagenesis. Four genes encoding para-nitrobenzyl (pNB) esterase variants exhibiting enhanced activity were recombined in two cycles of high-fidelity DNA shuffling and screening. Genes encoding enzymes exhibiting further improvements in activity were analyzed in order to elucidate evolutionary processes at the DNA level and begin to provide an experimental basis for choosing *in vitro* evolution strategies and setting key parameters for recombination. DNA sequencing of improved variants from the two rounds of DNA shuffling confirmed important features of the recombination process: rapid fixation and accumulation of beneficial mutations from multiple parent sequences as well as removal of silent and deleterious mutations. The five to sixfold further enhancement of total activity towards the para-nitrophenyl (pNP) ester of loracarbef was obtained through recombination of mutations from several parent sequences as well as new point mutations. Computer simulations of recombination and screening illustrate the trade-offs between recombining fewer parent sequences (in order to reduce screening requirements) and lowering the potential for further evolution. Search strategies which may substantially reduce screening requirements in certain situations are described.

Keywords: directed evolution; DNA shuffling; random mutagenesis; para-nitrobenzyl esterase

Introduction

Enzymes can be evolved *in vitro* to exhibit new and useful functions. A sampling of the local sequence space of the enzyme is created by mutagenesis; screening or selection directs the evolution towards the desired features. A successful strategy for improving enzyme activity in non-natural environments (Chen & Arnold, 1993) and on non-natural substrates (Moore & Arnold, 1996) has been to accumulate amino acid substitutions over multiple generations of random mutagenesis and screening. In practice, the best variant identified in each generation is chosen to parent the subsequent generation. Other potentially useful variants are set aside, and their mutations must be rediscovered in the evolved protein background in order to become incorporated. Because there is no mechanism other than back mutation for deleting mutations, this approach can also accumulate deleterious mutations, leading to premature termination of an evolving lineage. These are the classical arguments for the benefits of recombination (sex) in evolution (Maynard Smith, 1988). Recombination allows more rapid accumulation of beneficial mutations present in a population. It also makes possible the removal of deleterious mutations which would otherwise accumulate in an asexual population, a phenomenon known to geneticists as Müller's ratchet (Müller, 1932). Recombination can

Present addresses: J. C. Moore, Merck & Co., Inc., P.O. Box 2000, RY80Y-110, Rahway, NJ 07065, USA; H.-M. Jin, SmithKline Beecham, Mail Code UE0548, 709 Swedeland Rd., King of Prussia, PA 19406, USA.

Abbreviations used: pNB, para-nitrobenzyl; pNP, para-nitrophenyl.

0022-2836/97/380336-12 $25.00/0/mb971252

provide similar benefits for *in vitro* molecular evolution (Stemmer, 1994a,b).

Bacillus subtilis *p*-nitrobenzyl (pNB) esterase catalyzes the hydrolysis of the para-nitrobenzyl esters of various cephalosporin-type antibiotics, a necessary step in their large-scale synthesis (Zock *et al.*, 1994). Using four generations of sequential random mutagenesis and screening, we evolved a series of pNB esterases up to 30 times more active towards hydrolysis of the pNB ester of loracarbef (LCN-pNB) in aqueous dimethylformamide (Moore & Arnold, 1996). During the fourth generation, a large number (~7500) of pNB esterase clones were screened and partially characterized in order to validate the rapid screening assay. Sixteen improved pNB esterase clones were identified, from which the five most active enzymes (>50% enhancements in activity over the parent enzyme) were characterized. DNA sequencing revealed four unique pNB esterases (Table 1). Due to the limitations of screening, evolved sequences are generated using a low rate of point mutagenesis and typically accumulate a single beneficial mutation per generation. A simple restriction/ligation experiment demonstrated that recombination of mutations present in at least two of those sequences could further improve pNB esterase activity. Recombining gene segments from two improved pNB esterase variants yielded an enzyme twice as active as the best parent. DNA sequencing demonstrated that mutations from each of the two parents were combined in the new sequence (I60V and L334S), while one neutral or slightly deleterious mutation was deleted (K267R; Moore & Arnold, 1996).

Stemmer recently introduced the technique of "DNA shuffling" to create novel genes by recombination of closely-related DNA sequences (Stemmer, 1994b). Because it also introduces new point mutations during reassembly of the DNA fragments, DNA shuffling alone has been effective for directed protein evolution starting from a single sequence (Stemmer, 1994a; Crameri *et al.*, 1996). Questions arise as to how this approach is best implemented and integrated with other *in vitro* evolution approaches such as sequential random mutagenesis. Issues include optimizing the point mutagenesis rate associated with DNA shuffling, determining appropriate screening sample sizes and how many parental genes to recombine, and deciding when to use recombination. Here we investigate the further evolution of pNB esterase by DNA shuffling of the improved sequences generated by random mutagenesis and screening. By following how the genes evolve during cycles of DNA shuffling and screening, we can elucidate the mechanisms contributing to the evolution of function and begin to optimize strategies for *in vitro* evolution. An analysis of the recombination process identifies some of its benefits and limitations for directed evolution and allows a rational choice of mutagenesis and screening strategies.

Results and Discussion

Recombination statistics and screening requirements

To comment on the utility of DNA shuffling in directed evolution, a review of the statistics of recombination of multiple parent sequences is useful. For this discussion, we will assume that the mutations are unique and distributed far enough from one another on the genes that recombination occurs freely between any two. Furthermore, equal amounts of the initial DNA sequences are recombined. Consider the random recombination of three parent sequences, each of which contains a single mutation. Any given mutation will be incorporated into a progeny sequence with a probability of 1/3; the probability of generating the wild-type sequence is 2/3 at each mutation site. This highlights an important consequence of shuffling multiple sequences: there is a statistical preference for the absence of mutation in the progeny. The overall probability of picking a completely wild-type sequence from the recombined library is $(2/3)^3 = 0.30$. The probability of generating a sequence containing a single mutation (a parent sequence) is $1/3 \times (2/3)^2 = 0.15$. Because there are $C_1^3 = 3!/1!2!$, or three such sequences, the overall fraction of parent sequences in the library is 0.45. Thus fully 75% of the sequences in the recombined library are variants already in the evolutionist's possession.

In general, for a recombination system consisting of N sequences and M total mutations, the probability of generating progeny sequences containing μ mutations equals the number of ways a μ-mutation sequence can be generated (C_μ^M) multiplied by the probability of generating any single μ-mutation sequence:

$$P_\mu = C_\mu^M \left(\frac{1}{N}\right)^\mu \left(\frac{N-1}{N}\right)^{M-\mu}$$

$$= \frac{M!}{(M-\mu)!\mu!} \left(\frac{1}{N}\right)^\mu \left(\frac{N-1}{N}\right)^{M-\mu}$$

Figure 1 summarizes the analysis for recombination of single-mutation parent sequences ($N = M$). The probability that recombination will return the zero-mutation "grandparent" or single-mutation parent sequences remains constant between 73 and 75%; only ~25% of the clones screened have sequences that have not already been examined. The probability of creating individual sequences declines dramatically with increasing numbers of parents. The least frequent sequences are those containing the majority of mutations from the parent population, and the sequence containing all the mutations ($\mu = M$) is of course the rarest. The probability P_M of generating the rarest sequence is $1/N^M$.

Because we are interested in the evolution of function, we need consider only those mutations responsible for functional differences among pro-

Table 1. DNA and amino acid substitutions in fourth, fifth and sixth generation evolved pNB esterases

Mutation	Amino acid substitution	4-54B9	4-38B9	4-53D5	4-43E7	5-6C8	5-5E4	5-4H4	5-4G2	5-4D12	5-2D3	6-10F1	6-1D12	6-1C7	6-1A6
ATC 27 → ATA									YES	yes					
CCT 33 → CCC		yes													
ATT 60 → GTT	I60V		yes			yes	yes		yes	yes	yes	yes	yes	yes	yes
GAT 81 → GAC															yes
TAT 84 → TAC				yes		yes									
AGT 94 → GGT	S94G				yes										
GGA 127 → GGG															yes
TCG 148 → TCC						yes									
TTT 149 → TTC									yes	yes					
GGC 227 → GGG	A227G										yes				
ATT 239 → ATC				yes			yes								
AGA 246 → AGG															yes
GCG 252 → GCT							yes								
AAA 267 → AGA	K267R	yes													
CCG 317 → TCG	P317S					yes						yes	yes	yes	yes
TTA 334 → GTA	L334V				yes										
TTA 334 → TCA	L334S	yes				yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
GCT 343 → GTT	A343V			yes				yes							yes
CAT 356 → CGT	H356R								yes	yes		yes		yes	
ACT 359 → GCT	T359A												yes		
ATT 464 → GTT	I464V								yes	yes		yes		yes	

DNA substitutions are identified in the context of the three-base codon of the encoded enzyme sequence. Grey background indicates mutations from fourth generation parent sequences. White background indicates new mutations which arose during DNA shuffling. Bold face type indicates translated DNA mutations.

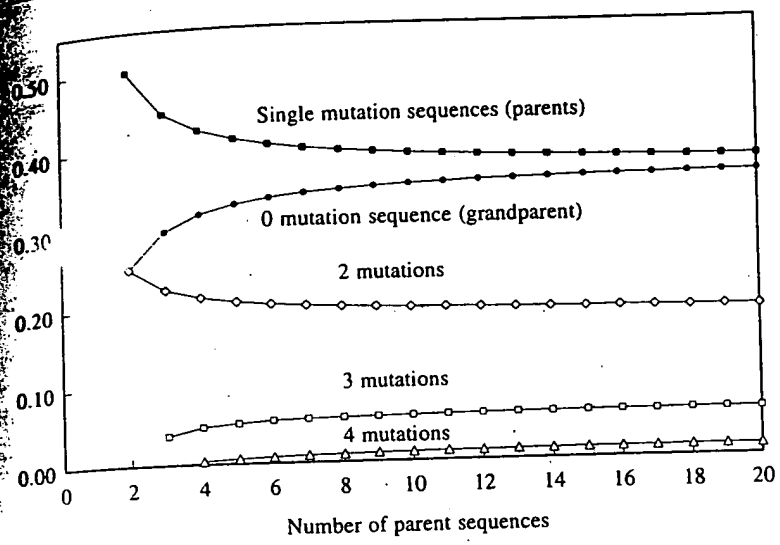

Figure 1. Probabilities of generating sequences containing different numbers of mutations by random recombination, based on recombining single-mutation parent sequences. Novel variants (not grandparent or parent sequences) are shown with unfilled symbols.

tein variants. Neutral mutations by definition do not affect function; their distribution among progeny sequences is determined statistically, even in the screened population (Zhao & Arnold, 1997b). Thus for the purposes of this discussion of recombination libraries and screening requirements, M is the number of mutations that affect the targeted function (either beneficial or deleterious).† By screening enough clones to ensure that the rarest sequence, that is, containing all M mutations, has been examined, one can be sure that the best variant will be discovered. This is true even if the best variant does not contain all the functional mutations (as would be expected if some mutations were deleterious or if the effects of mutations are not cumulative).

In practice, of course, oversampling is required to ensure that a particular variant has been examined during the course of screening. To be 95% confident that the most active combination variant has been examined, we must be 95% confident the rarest variant has been examined. If S is the number of clones sampled, then

$$(1 - P_M)^S < 1 - \text{confidence limit}$$

describes how the probability of not sampling the rarest variant changes with increasing S. This allows calculation of the number of samples required for a given confidence limit. The oversampling is then how many more samples must be screened over the theoretical minimum. When one clone is required with 95% confidence, the oversampling will be between 2.6 and 3.0 (for larger numbers of parents). Even a relatively low rate of background point mutagenesis, however, can introduce significant confounding effects. Non-neutral point mutations obscure recombination events and increase the amount of screening required to find the best sequences (*vide infra*). Thus, in practice, it may be impossible to screen sufficient numbers of clones to be sure of finding the best recombinant, particularly when the point mutation rate is high and a large number of functional mutations are being recombined. Alternative strategies which can reduce screening requirements under special conditions will be discussed further on.

DNA shuffling of evolved pNB esterases

An effect of forcing DNA polymerase to synthesize full length genes from the pool of small DNA fragments generated during DNA shuffling is additional background point mutagenesis. A high rate of point mutagenesis can severely inhibit the discovery of novel combinations of existing mutations within a population. Because most mutations are deleterious (in a screening assay sensitive to small changes in the screening variable), beneficial recombinations and rare beneficial point mutations are masked by the negative background. DNA shuffling with a 0.7% mutagenesis rate, for example, would yield an average of 10-11 point mutations in the 1470 bp pNB esterase gene. This is substantially more than the optimal mutation frequency (~three mutations per gene) for directed evolution of pNB esterase (Moore & Arnold, 1996). In fact, when the four evolved pNB esterase gene sequences were shuffled using *Taq* polymerase, fully 90% of the clones in the resulting library exhibited essentially no esterase activity during screening (data not shown). In a parallel study, we observed that 80% of the clones generated by DNA shuffling of subtilisin E exhibited no activity (Zhao & Arnold, 1997a).

In an effort to reduce the background mutagenesis rate, a proofreading polymerase (Pwo) was used during fragment reassembly. With Pwo, 50 to 100 base-pair fragments could be reassembled to create a library in which fully 80% of the clones

† A mutation that is neutral in one context (i.e. in the wild-type background), but becomes functional in a different context, would be considered a functional mutation.

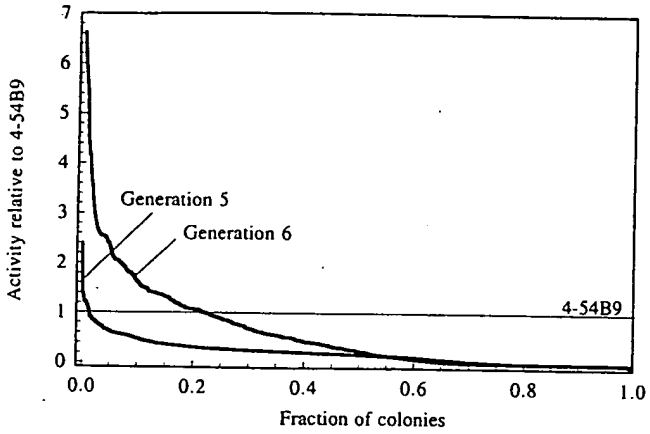

Figure 2. Activity profiles of generations 5 and 6 determined by screening libraries created by DNA shuffling of unique fourth and fifth generation variants. Activities were sorted from best to worst. Profiles are normalized by the number of clones screened.

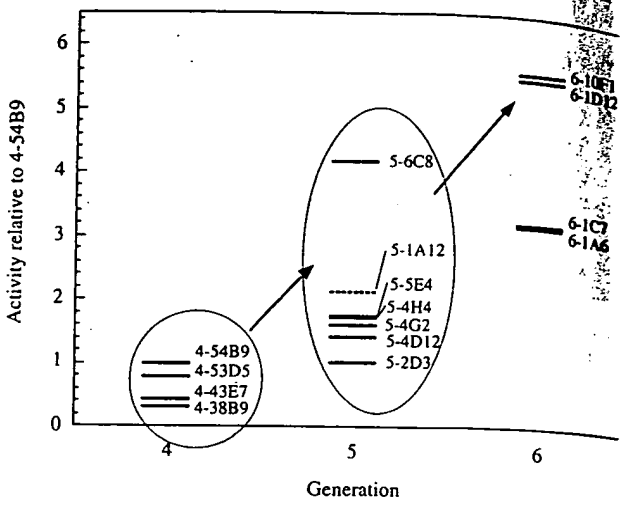

Figure 3. Activities of fourth, fifth and sixth generation pNB esterase variants (Table 1) in screening assay. Fourth generation variants were recombined and screened to identify improved enzymes in generations 5 and 6.

retained activity. Inserts from 13 randomly picked colonies were partially sequenced in order to determine the point mutation rate. Five mutations not present in any of the parent sequences were found in 12,000 nucleotides sequenced, for an overall mutagenic rate of ~0.04%. These minimally mutagenic conditions were used for DNA shuffling. A subsequent, in-depth investigation of the various steps involved in DNA shuffling has allowed us to identify a set of recombination protocols with a wide range of point mutagenesis rates (Zhao & Arnold, 1997a).

Four unique fourth generation improved pNB esterase variants were chosen as the starting point for further directed evolution by DNA shuffling. Two cycles of DNA shuffling and screening for activity towards the *p*-nitrophenyl ester of loracarbef (pNP-LCN) in 25% dimethylformamide (DMF) were performed. The activity profiles of the resulting populations (generations 5 and 6) are shown in Figure 2. To generate these profiles, activities of the individual clones measured in the 96-well plate screening assay were normalized by cell density (A_{600}) and plotted in descending order. Approximately 2% of the 948 generation 5 clones screened exhibit more total activity than the most active parent (4-54B9). The screened population was sufficiently large to give a high level of confidence that the most active variant that can be generated by simple recombination of the fourth generation sequences has been found.† The six most active variants from generation 5 were collected and shuffled again to create generation 6. Fully 20% of the 474 clones screened were more active than 4-54B9. Only 20 to 25% of the clones were inactive, as expected using the high fidelity Pwo-only shuffling conditions.

Figure 3 summarizes the activities of the four fourth generation parents and the best variants identified in generations 5 and 6. The improvement in enzyme activity as a result of shuffling is already apparent in the fifth generation, which includes one variant (5-6C8) fourfold more active than 4-54B9 and twice as active as variant 5-1A12 previously generated by ligation recombination (Moore & Arnold, 1996). The sixth generation contains two clones with yet higher activities than 5-6C8. The best one, 6-10F1, represents a five to six-fold improvement over 4-54B9 and is ~150 times more active than the wild-type.

Activities of the fifth and sixth generation variants towards the *p*-nitrobenzyl ester of loracarbef (LCN-pNB) were also determined, using a modified HPLC assay as described in Materials and Methods. The best pNB esterase is 5-6C8, which exhibits a threefold increase in total activity over 4-54B9. This clone is now ~100 times more active than wild-type pNB esterase towards LCN-pNB in 25% DMF. The sixth generation variants exhibited no further improvement in activity towards this substrate, a clear reflection of the use of the pNP ester during screening and the first law of random mutagenesis: "You get what you screen for" (You & Arnold, 1996).

† When shuffling four parent sequences each of which contains one beneficial mutation, 765 clones must be screened to be 95% confident that all combinations have been examined (assuming recombination occurs freely between mutations and no point mutagenesis). A 0.04% rate of point mutagenesis translates to less than 0.6 new mutations per sequence, of which only a fraction will affect function (estimated from the activity profile of a library created by error-prone PCR to be ~0.5, data not shown).

Analysis of evolved pNB esterase genes

DNA mutations present in the four parent fourth generation sequences and mutations identified by sequencing the genes encoding the selected fifth and sixth generation variants are summarized in Table 1. By comparing the activities and sequences of these variants with the third-generation parent, four beneficial mutations were identified (leading to amino acid substitutions I60V, L334V, L334S and A343V). The remaining mutations present in the fourth generation sequences are neutral or mildly deleterious (Moore & Arnold, 1996).

Several interesting observations can be made from this Table. It can be seen that a number of mutations increase their frequencies in the subsequent generations. Substitutions I60V in 4-38B9 and L334S in 4-54B9 are each present in a single fourth generation parent. In contrast, I60V is present in five of the six fifth-generation variants, and L334S is present in all six. By the sixth generation both substitutions are fixed in the population. A new substitution at position 317, first found during the fifth generation (5-6C8), also becomes fixed by the sixth. This new mutation probably accounts for the significant increase in activity of variant 5-6C8. The P317S substitution is positioned near the enzyme surface in a loop located on the same side of the entrance to the substrate binding pocket as amino acid substitutions L334S, M358V and A343V (Moore & Arnold, 1996). Removal of a proline at this position may relax conformational constraints on the loop, allowing the substrate freer access to the active site.

The two separate beneficial mutations at position 334 in 4-43E7 and 4-54B9 are mutually exclusive, and a competition exists as to which one will be propagated to successive generations. Variant 4-54B9 has more than twice the activity of 4-43E7 as a result of the mutation at position 334, and the fifth generation recombination progeny in fact show the L334S substitution from 4-54B9 exclusively. Recombination provides a rapid means to identify the most effective mutation among multiple possibilities at any given site.

Related to the observation that beneficial mutation combinations are fixed is the fact that recombination and screening also effectively remove neutral and deleterious mutations. Three of the five mutations present in the fourth generation parents that are synonymous (DNA mutations in codons 33, 84, and 239 that do not lead to amino acid substitutions) or non-synonymous, but believed neutral or mildly deleterious in their effects on total activity (mutations leading to amino acid substitutions S94G and K267R (Moore & Arnold, 1996)), have been removed from the improved pNB esterase population in a single round of shuffling; all five are removed by the sixth generation. The two most active sixth generation enzyme variants, 6-10F1 and 6-1D12, have no synonymous mutations at all and only one mutation (at position 359) not seen in any previous clone. Due to the statistical preference for the absence of mutations the recombination process is highly effective in filtering out neutral (and deleterious) mutations starting from multiple parent sequences.

Table 1 also shows that the DNA shuffling technique can recombine multiple parent sequences to create novel progeny. Recombination between at least three fourth-generation parents is required to create 5-5E4, and at least three fifth-generation parents were recombined to generate clones 6-10F1 and 6-1A6 (based on the presence and absence of the DNA mutations in the sequences compared to the parent sequences).

Finally, it is useful to note that DNA shuffling generates point mutations that are rarely observed during PCR (at least for the low-mutagenesis rate PCR conditions used for directed evolution of longer DNA sequences). Four of the 12 new point mutations identified in the fifth and sixth generation variants, for example, are G → C (and C → G) and G → T (and C → A) transversions, which were not found at all during the first four generations of pNB esterase evolution involving PCR mutagenesis (Moore & Arnold, 1996). These mutations were also generated very rarely during the error-prone PCR mutagenesis of subtilisin (Shafikhani *et al.*, 1997). DNA shuffling and error-prone PCR together may provide access to a wider range of amino acid substitutions.

Evolved pNB esterase amino acid sequences

Amino acid substitutions in the evolved pNB esterases are indicated in Table 1; changes in amino acid sequence along the lineage are summarized in Figure 4. The accumulation and fixation of two beneficial amino acid substitutions from the fourth generation, I60V and L334S, is essentially complete in a single generation of DNA shuffling and screening 948 clones. In contrast, A343V, a beneficial mutation found in the fourth generation, no longer appears in the majority of fifth or sixth generation variants. The (5-4H4) recombinant of the parent containing this mutation (4-53D5) with 4-54B9 shows no improvement in activity over 4-54B9 (Figure 3). Substitutions A343V and L334S therefore do not work in concert to improve enzyme activity, and consequently there is little or no driving force to retain A343V in the population. The remaining fifth generation variants, with the exception of 5-6C8, are less active than 5-1A12 (Figure 3), yet they contain the I60V and L334S substitutions while omitting K267R, as does 5-1A12. This suggests that the additional mutations found in those sequences are neutral, or possibly, deleterious. For instance, the amino acid sequences of 5-5E4 and 5-1A12 are identical, and the decreased activity of the former is likely due to the two synonymous mutations in 5-5E4 not present in 5-1A12. Because the screen evaluates the total activity of a clone (normalized by cell density), synonymous mutations can influence the result, for

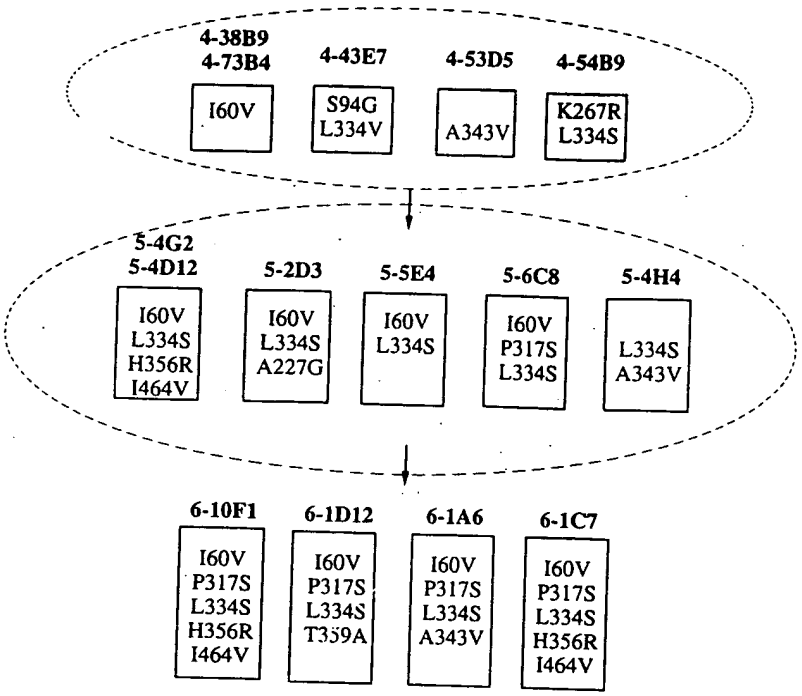

Figure 4. Lineage of pNB esterase variants showing amino acid substitutions accumulated by four generations of sequential random mutagenesis (fourth generation) and by DNA shuffling (fifth and sixth generations) and screening. All variants contain amino acid substitutions H322R, Y370F, M358V and L144M from the third generation parent (Moore & Arnold, 1996).

example, by affecting the amount of active enzyme expressed. The new beneficial mutation that gives rise to the P317S substitution becomes fixed in the sixth generation, and further evolution during that generation primarily arises from point mutation rather than recombination.

Clones 5-4G2 and 5-4D12, whose DNA sequences are identical, both contain amino acid substitutions H356R and I464V. These two substitutions are seen together again in 6-10F1 and 6-1C7. Because 6-10F1 and 6-1D12 have almost identical activity, we can reasonably infer that the I60V, P317S, and L334S substitutions are responsible for that activity, while the mutations leading to H356R and I464V from the fifth generation as well as a new mutation, T359A, in 6-1D12 are neutral. The three mutations believed responsible for enhanced activity are also present in 6-1A6, along with the last mutation in this system known to enhance activity, A343V. That 6-1A6 has lower activity than 6-10F1 and 6-1D12 is therefore attributable to either the three synonymous mutations in 6-1A6 (Table 1) or antagonism between amino acid substitutions A343V and P317S or I60V.

The new point mutations that arose during the minimally mutagenic DNA shuffling increased (P317S) and decreased enzyme activity. The effects of individual mutations can be ascertained with confidence because the sequences differ from one another at very few positions. We have recently demonstrated a method that allows one to distinguish clearly beneficial, neutral and deleterious mutations in evolved sequences by random recombination with ancestor sequences (Zhao & Arnold, 1997b). This method will be particularly useful for identifying mutations responsible for functional changes in proteins in a background of neutral mutations (as happens when multiple new mutations are present).

Only 2% of the fifth generation clones are more active than the most active parent, 4-54B9 (Figure 2). Although 25% of the progeny should be novel, the combination I60V + L334S predominates in the most active variants (Figure 4), suggesting that many of the remaining combinations lead to lower activity than in 4-54B9. Additionally, while there is no mechanism for recombination alone to generate inactive clones, ~25% of the variants in Figure 2 are inactive, presumably as a result of background point mutation. This implies that the frequency of enhanced-activity recombinants is reduced by point mutation and emphasizes the importance of minimizing the mutagenesis rate when recombining positive mutations.

Developing strategies for directed evolution

Recombination versus *random mutagenesis*

Recombination is only useful if a population of sequences is available from which new combinations of mutations can be generated. Homologous proteins with similar sequences could provide such a starting population (Stemmer, 1994b). (Note, however, that a high level of sequence identity may be required for DNA shuffling.) Populations of sequences can also be created by the background point mutagenesis feature of DNA shuffling (Crameri *et al.*, 1996). Alternatively, they can be generated by random mutagenesis and screening experiments, as they have been for the current study. When interesting sequences already exist, recombination offers an efficient means to use that information. If the sequences must be generated, however, then one should consider that

...in the overall cost of evolution by recombination as compared to, for example, evolution by ...uential generations of random mutagenesis and ...ening.

In theory, the sequential (or "asexual") approach ...uiring the least labor in terms of screening is to ...een randomly mutagenized clones until a positive is ...ified and then use that as the template ...r the next generation. The process is a random ...alk in which the first uphill step encountered is ...ken. To take a simple illustration, consider three ...utations A, B and C that each contribute in a ...umulative, if not additive, manner when combined. A, B and C could be collected in the ABC ...variant in three sequential generations of mutagenesis and screening. Alternatively, if A, B and C all contribute to the desired feature in the wild-type back...d (as they often do; see, for example, Chen & Arnold, 1993), they could be found separately and then recombined to make ABC. Finding the single-mutation sequences A, B, and C, however, requires screening the same number of colonies as finding ABC by sequential evolution. Recombining the A, B, and C sequences to make ABC requires additional screening. Of course, the sequential pathway requires three random mutagenesis steps, while the recombination pathway re... only one mutagenesis step and one DNA shuffling step. The advantages of one approach over the other then depend on the costs of screening relative to the DNA manipulations.

Note that the severe limitations screening places on the number of colonies that can be sampled makes it difficult to accept downhill steps in the hope that further improvements can be found further out in sequence space (Moore & Arnold, 1995). It also means that extremely rare events ... as the recombination of neutral or slightly deleterious mutations to make a beneficial combination will probably not contribute in any significant fashion to the evolutionary process.

The pNB esterase evolution provides a concrete example for analysis. Approximately one in every 1500 to 2000 randomly mutagenized pNB esterase clones screened was positive (showing 50% or greater enhancement in activity over the parent; ...oore & Arnold, 1996). To generate the population ... four unique positives for DNA shuffling, we examined a total of 7500 clones. Finding the best combination variant required additional DNA shuffling experiments, and ~1400 additional colonies were screened. Thus a total of 9000 clones were screened in going from generations 3 to 6. There is no guarantee that the sequences chosen for recombination are unique: in fact, the original fourth generation clones contained five variants, ...wo of which were identical (4-38B9 and 4-54B9) ...nd two of which contained mutations in the same codon (4-43E7 and 4-54B9), precluding recombination between these variant pairs. It is very likely that variants of comparable or even greater activity could also have been created by continuing random mutagenesis and screening for three generations from the first fourth generation variant identified. The total screening requirement would be the same.

In practice, however, the uphill climb often involves identification of multiple positives during each generation. Everything but the one chosen to parent the next generation is discarded in the random uphill walk of the "asexual" evolution. During the pNB esterase evolution, we often identified four or five potential positives during the rapid screen on the LCN-pNP colorimetric substrate. Those were either verified or not during a second level screen on the *p*-nitrobenzyl (LCN-pNB) substrate, and it was often the case that more than one sequence was a true positive (Moore & Arnold, 1996). The other improved sequences could of course be collected and recombined at any time and at relatively little screening cost. A significant advantage of the DNA shuffling method is its ability to utilize these available positive sequences.

Computer simulations of random recombination and screening

The statistical model can be used to optimize the number of parent sequences chosen for DNA shuffling. Screening during the fourth generation actually resulted in the identification of 16 clones measurably more active than the parent, of which five were at least 50% more active (Moore & Arnold, 1996). An attempt to recombine all 16 sequences yielded no clones more active than 4-54B9 (~1000 clones screened). This result can be understood when we consider the dramatically lower probability of finding the best combination(s) as the number of sequences increases. If the screening sample size is limited to a few thousand clones, there is little chance that the best sequences, or even sequences better than the best parent, will be found by screening a library created from 16 parents.

We have used a computer simulation of the random sampling of the two recombined libraries obtained by shuffling five and ten sequences to illustrate the advantage of choosing fewer parents when screening is limited. Recombining all ten parents becomes advantageous, however, when large numbers of clones can be examined. (Of course, the larger sampling requirement should then be compared to the potential for continued evolution by random mutagenesis.) Assuming that the ten parent sequences each contain a unique, single beneficial mutation ($N = M$) and that they can be recombined to give all possible combinations, we calculated P_μ for $\mu = 0$ through 10. Since $\Sigma P_\mu = 1$, these were organized into a cumulative distribution from 0 to 1, and a random number generator was used to pick a point on the cumulative distribution, thereby identifying μ (number of mutations per sequence). A second random number generator was used to pick one of the C_μ^M possible sequences containing μ substitutions using an evenly spaced distribution of possible

combinations. The activity of the sequence chosen was then calculated by assuming that the free energies of activation of the variants (proportional to the natural logarithms of their activities) are additive.

The results of this simulation are shown in Figure 5, using the activity data from the fourth generation pNB esterase variants. Figure 5(a) shows the averages of the highest values of mutant activities obtained over 15 separate trials for each (screening) sample size. The results obtained by shuffling the ten best mutants (black diamonds) can be seen to be slightly worse than those obtained by shuffling the five best mutants (white squares), for sample sizes up to about 10,000 to 15,000. That is, the average expected best mutant is higher for shuffling five parents at a time for small sample sizes. Figure 5(b) and (c) show the range of values of the highest mutant activity obtained on each of 15 separate trials for each sample size. Here, the highest values obtained from recombining the best ten variants (black diamonds) become better than the values obtained from shuffling the best five (white squares) at sample sizes greater than about 1000. Although shuffling the top ten mutants for this set of data can yield higher final activities, the simulation shows that the outcome is much more risky when screening capabilities are limited to a few thousand clones.

Simulations also show that the results of the comparison of shuffling five *versus* ten parents is highly sensitive to the values of the activities. For instance, if the activities of mutants 6 through 10 are decreased, then the sample size at which recombining all ten mutants becomes preferable becomes much higher. Moreover, the simulation can be adapted for cases in which some or all of the parent sequences have two or more mutations, which may or may not be recombinable. Thus this simulation approach can be used to determine the optimal number of sequences to recombine for any given set of activity values and any given sample size.

The simple additivity assumption on which these simulations are based† is a reasonable first approximation of the behavior of combined mutations in proteins (Wells, 1990) and is useful for a first exploration of strategic issues in *in vitro* protein evolution. The real behavior is often more complex and will depend on the property of interest as well as the particular protein. However, it is likely that deviations from simple additivity are governed by non-linear functions of the number and magnitude of changes; values will certainly depend on which subset of mutations is recombined. While it is possible to modify the simulation to take into account deviations from additivity, very little data are available on the effects of large numbers of mutations. We have therefore not

† Both beneficial and deleterious mutations can be accommodated in this framework.

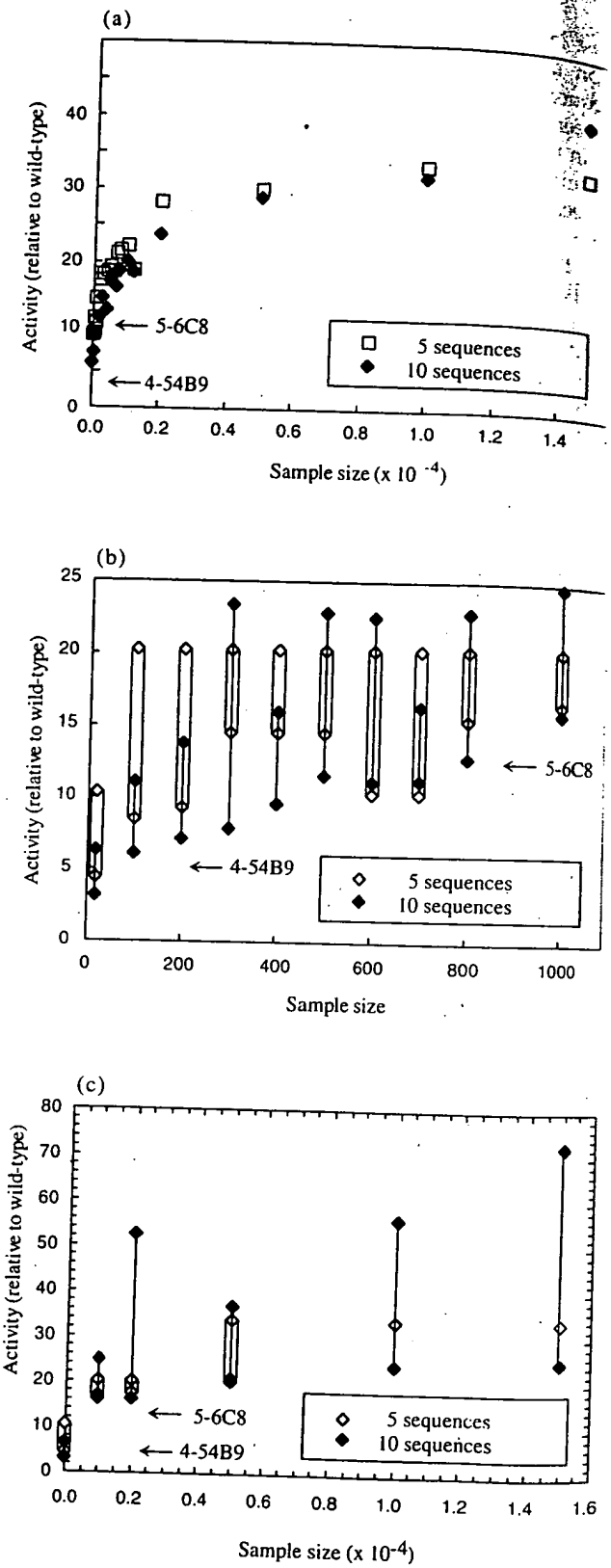

Figure 5. (a) Averages of highest values of mutant activities obtained over 15 separate trials of simulated random recombination of five and ten parent sequences. (b) and (c) Range of values of mutant activities obtained over 15 separate trials. Activities of best fourth-generation parent (4-54B9) and highest-activity fifth generation clone identified (5-6C8) are indicated for comparison.

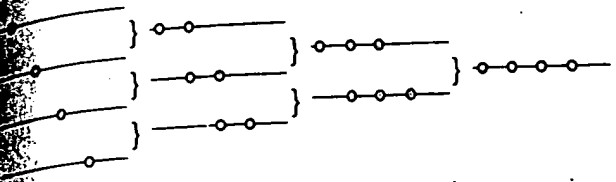

Figure 4. Pairwise recombination can reduce screening requirements, provided effects of mutations are cumulative. By shuffling two sequences at a time, sequences containing two mutations represent 25% of the recombined library. This example involves six recombination experiments.

attempted to include deviations from additivity in the current simulations. Figures 5(a), (b), and (c) show the activities of the best fourth generation parent (4-54B9) and the best fifth generation clone identified (5-6C8) by screening the shuffled library. That the activity of 5-6C8 is ~twofold less than the average expected for screening 948 clones reflects the fact that (i) only four of the original five positive clones identified during generation 4 were unique, (ii) two mutations were on the same codon and could not be recombined, and (iii) the mutations combine with significantly less than linear additivity.

Alternative search strategies

Finally, we will briefly consider two other search strategies that might be used to minimize screening requirements. One approach to producing a multiple-mutation variant which requires the screening of less clones is multiple-step pairwise recombination. This strategy is illustrated in Figure 6 for the simple case of recombining four (beneficial) mutations from four separate parents. Pairs of parents are mated. As each progeny is a double mutant 25% of the time, only 12 clones are required to find all the double mutants, assuming the effects of the mutations are cumulative. The double mutants are then similarly mated, and screening only eight clones will identify the triple mutants. Mating and screening four clones will generate the quadruple mutant. Thus a total of only 62 clones (24 × 2.6 times oversampling to be 95% confident at each step) must be screened, as compared to the 765 required to generate the quadruple mutant in a single recombination step. Such an approach requires considerable DNA manipulation and would be most useful when screening is extremely difficult. (An attractive alternative at this point may be sequencing the parents and recombination by site-directed mutagenesis.) A further cost of this approach is that the search space is very limited. The assumption is that each activity-enhancing mutation will contribute to the overall activity, so that the quadruple mutant is the best performer of this population. If a particular double or triple mutant is the best performer, it may or may not be found, since not all of these intermediate mutants will have been examined.

A compromise method that works well, at least in theory, can be described as "population recombination." The idea is to shuffle all four parent sequences at once and screen enough clones to see all the double mutants. Because each double mutant occurs 3.5% of the time, 28 clones must be screened. This examines all of the pair-wise interactions between mutations and eliminates those which are not cumulative. The double mutant population is recombined to produce all of the triple mutants and the quadruple mutant (requires screening 16 clones). If the mutations were at least cumulative in their effects, screening 132 (44 × 3.0 times oversampling) clones would search the space completely for the best (quadruple) mutant. This approach most closely describes how recombination/selection experiments operate (Stemmer, 1994a) where all of the clones that survive a particular selection criterion are recombined (often 100 clones or more serving as the parent population for the next generation).

Conclusions

Recombination is an important tool for directing the evolution of proteins. Beneficial mutations can be recombined, while neutral and deleterious mutations are eliminated. The need to screen rather than select for many important enzyme functions, however, severely limits the ability to search for useful combinations. It is therefore imperative to analyze various recombination strategies. Mutagenic rates associated with the recombination process must be low so that beneficial mutations are not lost in a background of deleterious ones. Although a new beneficial amino acid substitution was found as a result of the DNA shuffling of pNB esterase, DNA shuffling may be less efficient for discovery of new mutations compared to a controlled mutagenesis technique (a beneficial mutation can be masked in the background of recombined sequences). Utilizing more than two parents for recombination introduces a statistical preference for not incorporating mutations in progeny, and this has several consequences especially with respect to screening. Recombination favors the dilution of progeny containing the most mutations, which has the effect of exponentially increasing the number of progeny that must be screened in order to find the rarest ones. Because shuffling large numbers of parent sequences can yield many possible combinations, it may also be necessary to strictly limit the number of parent sequences in any given recombination experiment. We have described two alternative search strategies which reduce the required number of variants examined, at the cost of possibly missing intermediate beneficial combinations.

Finally, recombination requires a population of positive variants for efficient enzyme improve-

ment. If a population of positive variants must first be generated, sequential random mutagenesis may require less effort to produce sequences containing multiple mutations. Multiple positive variants are often generated, however, during a single cycle of random mutagenesis and screening. Recombination of these positives can provide substantial improvements at relatively little cost.

Materials and Methods

DNA shuffling

DNA shuffling was performed as described by Stemmer (1994b) with modifications. The 2 kb DNA fragment encoding the *B. subtilis* pNB esterase gene was amplified using PCR (forward primer 5'-CAATCTAGAGGGTATTAATAATG-3' and reverse primer 5'-CGCGGGATCCCCGGGTACCGGGC-3'). The amplified DNA was purified by gel electrophoresis and extraction using Qiaex kit (Qiagen, Chatsworth, CA). A total quantity of ~10 µg DNA, either from a single parent (non-recombinatorial) or from a mixture of multiple parent sequences (recombinatorial), was digested with DNase I (0.0015 units/µl) at room temperature for 20 minutes in a 100 µl reaction. After ethanol precipitation, the digested DNA was electrophoresed as a smear in a 3% low melting temperature gel of NuSieve GTG Agarose (FMC Bio Products, Rockland, ME). DNA fragments in specified molecular size ranges were collected onto DE-81 filter paper disks (Whatman, Maidstone, England) by electrophoresis and eluted from the filter paper with 400 µl of 10 mM Tris/1 mM EDTA buffer (pH 8.0) containing 1 M NaCl. The DNA fragments were ethanol precipitated and redissolved to approximately 20 ng DNA/µl in 1 × Pwo DNA polymerase buffer (Boerhinger Mannheim, Indianapolis, IN) containing 2 mM $MgSO_4$ and 0.2 mM each of the four dNTPs. A 5 unit/µl Pwo DNA polymerase solution (Boehringer Mannheim) was diluted tenfold, and 0.5 µl was added to a 5 µl redissolved DNA reaction solution. Reassembly of DNA fragments was conducted by PCR, using the conditions 94°C for 40 seconds., then 70 cycles of 94°C for 30 seconds, 50°C for 30 seconds, 72°C for 30 seconds, followed by a final extension step at 72°C for five minutes. A second 0.5 µl of Pwo polymerase was added at the annealing step of the 35th cycle. The reassembled DNA fragments were amplified in a conventional PCR (25 cycles) with the dilution of 1 µl reassembled DNA fragments in a 100 µl reaction. Once the success of the reassembly/amplification reactions was verified by gel electrophoresis, the reassembled product was purified with a Wizard PCR prep kit (Promega Corp., Madison, WI), digested with *Bam*HI and *Xba*I, concentrated by ethanol precipitation, and electrophoresed in an agarose gel. The 1.8 kb product was cut from the gel and the DNA extracted using a Qiaex kit. The final products were ligated with the vector generated by *Bam*HI-*Xba*I digestion of pNB106R (Zock *et al.*, 1994). This library was used to transform competent *E. coli* TG1 cells, as described (Moore & Arnold, 1996).

Screening a pNB esterase library

Screening was based on the assay described previously (Moore & Arnold, 1996), using the *p*-nitrophenyl ester of the loracarbef nucleus (LCN-pNP) as substra *E. coli* TG1 containing the plasmid library were gro on LB/tetracycline (20 µg/ml) plates. After 36 hours 30°C single colonies were picked into 96-well plates co taining 100 µl LB/tetracycline medium per well. The plates were shaken and incubated at 30°C for 12 hou to let the cells grow to saturation. Aliquots (20 µl) of t cultures were inoculated into a fresh plate containi 100 µl media per well; these were incubated at 40°C f ten hours with shaking to induce the expression of pN esterase. Esterase activities were then measured by trar ferring 20 µl aliquots of the cell cultures into a fresh s of plates where they were mixed with 200 µl of 0.1 Tris-HCl (pH 7.0) 25% DMF and 2 mM LCN-pNP. Rea tion velocities were measured at 450 nm over ten mi utes. (11 data points) in a ThermoMax microplate read (Molecular Devices, Sunnyvale CA). Activities were no malized by the cell densities of the original wel measured at 600 nm to control for variations in ce quantities.

For each round of screening, the clones that showe the highest activities were re-streaked on LB/tetracyclir agar plates, and single colonies derived from these plat (three to four colonies from each clone) were inoculate simultaneously into 96-well plates and tube cultures. Th former were used to repeat the activity assay, a described above, and the latter were used for glycer stock and plasmid preparation (Qiawell kit, Qiagen).

Assay of pNB esterase activity on LCN-pNB

A modified HPLC assay was used to determin enzyme activity towards the LCN-pNB (*p*-nitrobenzy ester) substrate (Chen *et al.*, 1995). The bacterial cell were incubated at 30°C with shaking for 12 hours an then at 40°C for ten hours to induce expression of pN esterase. Aliquots of cells (200 µl) were incubated wit 300 µl reaction buffer for 30 minutes at room tempera ture. The final reaction mixtures contained 0.1 M Tris HCl (pH 7.0) 25% DMF and 2 mM LCN-pNB. The reac tions were stopped by addition of 500 µl acetonitrile an passed through a nylon syringe filter (Micron Separations, Inc., Westboro, MA) with a pore size of 0.45 µm. Aliquots of each sample (50 µl) were analyzed by HPLC on a 250 mm × 4.6 mm C18 reverse-phase colum (Vydac, Hesperia, CA) at room temperature using a linear gradient starting with 50:50 of A:B (A = 5% methanol/95% 1 mM triethylamine, pH 2.5; B = 100% methanol) and ending with pure B in eight minutes (flow rate of 1 ml per minutes). Product and substrate were detected at 270 nm. The area of the *p*-nitrobenzyl alcohol product peak was calculated and subtracted from the area of the same peak from a sample containing *E. coli* without a pNB esterase gene. This controls for the small quantities of free product in the substrate preparation and any interference from bacterial contamination. This final area was used as a measure of total activity, which was normalized by cell density.

Acknowledgments

The authors thank Dr W. P. C. Stemmer (Maxygen) for many helpful discussions, Dr Steve Queener (Eli Lilly & Co.) for providing us the wild-type pNB esterase and the challenge to evolve it, and Ms Rebecca Little (Eli Lilly & Co.) for assistance with DNA sequencing. This

search is supported by the US Department of Energy's ogram in Biological and Chemical Technologies search within the Office of Industrial Technologies, ergy Efficiency and Renewables. O. K. is supported by NIH predoctoral training fellowship from the tional Institute of General Medical Sciences, NRSA ward 1 T32 GM 08346-01.

References

Chen, K. & Arnold, F. (1993). Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Proc. Natl Acad. Sci. USA*, **90**, 5618–5622.

Chen, Y., Usui, S., Queener, S. W. & Yu, C. (1995). Purification and properties of *p*-nitrobenzyl esterase from *Bacillus subtilis*. *J. Ind. Micro.* **15**, 10–18.

Crameri, A., Whitehorn, E. A., Tate, E. & Stemmer, W. P. C. (1996). Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nature Biotech.* **14**, 315–319.

Maynard Smith, J. (1988). The Evolution of Recombination. In *The Evolution Of Sex: An Examination Of Current Ideas*, pp. 106–125, Sinauer Associates, Inc, Sunderland, Mass.

Moore, J. C. & Arnold, F. H. (1996). Directed evolution of a para-nitrobenzyl esterase for aqueous-organic solvents. *Nature Biotech.* **14**, 458–467.

Moore, J. C. & Arnold, F. H. (1997). Optimization of industrial enzymes by directed evolution. *Advan. Biochem. Eng.* 58, 1–14.

Müller, H. J. (1932). Some genetic aspects of sex. *Amer. Nature*, **66**, 118–138.

Shafikhani, S., Siegel, R. A., Ferrari, E. & Schellenberger, V. (1997). Generation of large libraries of random mutants in *Bacillus subtilis* by PCR-based plasmid multimerization. *Biotechniques*, in the press.

Stemmer, W. P. C. (1994a). Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature*, **370**, 389–391.

Stemmer, W. P. C. (1994b). DNA shuffling by random fragmentation and reassembly: *in vitro* recombination for molecular evolution. *Proc. Natl Acad. Sci. USA*, **91**, 10747–10751.

Wells, J. A. (1990). Additivity of mutational effects in proteins. *Biochemistry*, **29**, 8509–8517.

You, L. & Arnold, F. H. (1996). Directed evolution of subtilisin E in *Bacillus subtilis* to enhance total activity in aqueous dimethylformamide. *Protein Eng.* **9**, 77–83.

Zhao, H. & Arnold, F. H. (1997a). Optimization of DNA shuffling for high fidelity recombination. *Nucl. Acids Res.* **25**, 1307–1308.

Zhao, H. & Arnold, F. H. (1997b). Functional and nonfunctional mutations distinguished by random recombination of homologous genes. *Proc. Natl Acad. Sci. USA*, **94**, 7997–8000.

Zock, J., Cantwell, C., Swartling, J., Hodges, R., Pohl, T., Sutton, K., Rosteck, P., Jr, McGilvray, D. & Queener, S. (1994). The *Bacillus subtilis* pnbA gene encoding *p*-nitrobenzyl esterase: cloning, sequence and high-level expression in *Escherichia coli*. *Gene*, **151**, 37–43.

Edited by J. Wells

(Received 10 March 1997; received in revised form 18 June 1997; accepted 9 July 1997)